

Is there a spin doctor in the house?

A report on science and society provides a useful overview of recent controversies and ways of approaching them. More could have been said about improving researchers' anticipation of the media and lobbyists.

On the one hand, there has never been a time when issues involving science have been more exciting, the public more interested, or the opportunities more apparent. On the other hand, public confidence in scientific advice to government has been rocked by a series of events. So says the select committee on science and technology of the upper house of Britain's parliament, the House of Lords. The committee's report, *Science and Society*, published this week, addresses a crisis of public trust in the United Kingdom in a detailed fashion that should have strong resonances elsewhere.

The committee raises a wealth of suggestions on rebuilding this trust (see <http://www.parliament.uk>), urging funding bodies to reward scientists who communicate their findings to the public, the government to support websites giving links to reliable sources of scientific information, and the Public Understanding of Science movement to adopt a less patronizing title. The interplay between science and the media is one of the report's strongest points. That is a relief, because recent pronouncements from the Lords' counterparts in the House of Commons have been naive and unhelpful.

Last May, the Commons science and technology committee reviewed the recent hostile media coverage of genetically modified (GM) foods. It recommended that "media coverage of scientific matters should be governed by a Code of Practice which stipulates that scientific stories should be factually accurate". This suggestion mystified many, given that an existing Press Complaints Commission code requires accuracy in all reporting. It smacked of special pleading for science — hardly likely to regain the public's trust.

Sensibly, the Lords committee rejects this call. "Science cannot

expect special treatment from the media," says its report. "Scientists must indeed take the rough with the smooth, and learn to work with the media as they are." In any case, the report notes, factual accuracy isn't the main issue: "Much more significant ... is the way in which the facts are used, both by writer and reader". In other words, 'spin'. The frenzy that gripped Britain last year was driven as much by public suspicion about the motives of large companies as by unease about biotechnology. The pressure groups that led the crusade against genetic modification recognized this, took some contentious results questioning the safety of GM food, and spun them for all they were worth. If scientific institutions are to respond effectively to such campaigns, they may need to hire their own spin doctors. Those in the firing line need better advice on how to present their arguments so that their own words aren't spun against them.

Good spin doctors would have deterred the government's chief scientific adviser from engaging in an ugly public row over GM coverage with the editor of a tabloid newspaper, as Sir Robert May did last year. They might also have warned the Royal Society about the way pressure groups would spin its criticism of Arpad Pusztai, the researcher whose unpublished findings sparked the GM scare. In addition to offering a scientific assessment of Pusztai's much publicized results, the society chastised him for talking to the media about research that hadn't been peer reviewed. Reasonable enough, but apply a little spin, and the story soon becomes "pro-GM scientific establishment gags whistleblower". The House of Lords committee says little about spin. But this is where important lessons from the GM experience can be learned if society's confidence in science is to be sustained. ■

A job half done

Spanish universities' need for research evaluation and competition is as strong as ever.

The newly re-elected Spanish government has greatly increased its majority. It would do well to focus on improving the assessment and related rewards for the country's universities. As a correspondent rightly suggests on page 222, a recent pioneering and bravely controversial evaluation of universities (see *Nature* 402, 848; 1999) left important questions unanswered, addressing teaching but doing a poor job in evaluating research. Unsurprisingly, low-scoring universities were the most critical whereas some of those with high scores were content to keep silent. Where next?

Will universities with a low score try to improve on their weak points, as listed in the study, in order to perform better in future and approach high-scoring universities? Probably not. The study held no consequences for the institutions it evaluated. Moreover, it has been attacked by public bodies responsible for assessing the quality of universities and also by rectors who got low scores, such as Saturnino de la Plaza, rector of the Polytechnic University of Madrid, who is also the

president of the influential Council of Rectors of the Spanish Universities. He says the council is to set up a commission that will evaluate the quality of universities but will not provide any type of ranking.

That approach will miss an important opportunity. Many countries rank universities or departments according to various measures of quality in research. Although Spanish students are being educated in an increasingly competitive system, open competition is still inadequate in important areas such as the recruitment of professors. Furthermore, a national evaluation committee is necessary to rank university departments on research and teaching, with more funding being awarded to the strongest. Research measures could include the volume and quality of publications and the number of patents. Moreover, to ensure a dynamic and competitive system, panels of foreign scientists should be invited to assess the quality of Spain's university research, as happens in Portugal. Such approaches are urgently required to help Spain deliver ever better teaching and research. ■

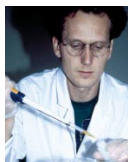




Irish fortune

Research gets a cash boost, but what should it be spent on?

p215



Slow DFG

German funding body 'takes too long to process grants'

p217



EMMA and EBI

EU research ministers say facilities should be saved

p217



Art at CERN

Particle physics has inspired a group of European artists

p219

Europe joins race to turn the Internet into one vast computer

Paris

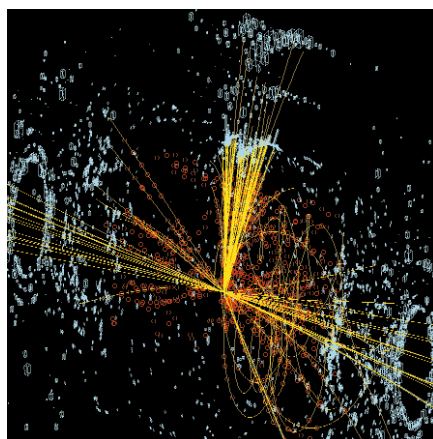
CERN, the birthplace of the World Wide Web, aims to make Europe a leading player in the largest distributed computing project in history. The Large Hadron Collider (LHC) at CERN, the European Laboratory for Particle Physics, will produce a deluge of data. Now CERN is coordinating a proposal that would use the challenge of analysing all these data to help develop a concept known as the grid.

The grid idea was created in the United States (see *Nature* **402**, suppl., C67–C70; 1999). Its goal is to develop software and Internet protocols to transform the Internet into a single gigantic computer. Researchers throughout the world could work on shared data sets on a network running thousands of times faster than today's best. Eventually, the technology could be applied to the public Internet.

The United States already has a \$500 million five-year grid effort involving 50 research centres and coordinated by the National Computational Science Alliance. CERN now plans to submit a proposal to the European Union (EU) for a grid infrastructure costing 30 million euros (US\$29 million). It has set itself a deadline of 10 May. The proposal is likely to be favourably received, as European Commission officials have actively solicited it.

David Williams, formerly CERN's head of computing and networks and now responsible for relations with the EU, points out that the web has made the physical location of information irrelevant, yet scientists still mostly use the Internet just to search for information. Real-time computing and data handling have hardly been explored, he says. John Taylor, director-general of the UK research councils and a convert to grids, has coined the term 'e-science' for this new way of working.

The CERN-led proposal will initially focus only on particle physics. But EU funding may be extended next year to other disciplines such as biology. Some European Commission officials see the challenge presented by the LHC's deluge of data as an opportunity for an all-out bid to catch up



Ring the changes: simulated data from the LHC.

with the United States in constructing what many believe will be the successor to the web.

Andrea Dahmen, spokeswoman for research commissioner Philippe Busquin, says he "wants to see a reliable high-performance EU Internet network in place as soon as possible".

The \$1.8 billion LHC, which will come online in 2005, is an ideal test bed for the grid concept. Collections of protons steered into head-on collisions will spew out around 7 petabytes (10^{15} bytes) of data every year. The raw data from just one of the LHC's detectors represent the equivalent of every person on Earth talking into 20 telephones at once. The challenge is to transmit this information — requiring 1,000 times more computing power than CERN can currently deliver — in a usable form to thousands of users in more than 40 countries.

Under the CERN plan, data would fan out across a high-speed network to 10 national and regional data centres, and in turn out to hundreds of local centres and universities. The innovation will be the glue holding all of this together: a new suite of middleware, software and protocols designed to allow real-time distributed computing. The Internet would be made to function as if it were a single computer and database rolled into one.

"Getting ourselves liberated from the geographical constraints will be crucially

important" for the success of the LHC, says Williams. "We need to use any and all possible resources to process the data. Not all of the people nor all of their computing and data-handling resources can be installed permanently at CERN."

Imagine an LHC scientist sitting at his or her desk in California, say, or Budapest. One click on a web wizard, and time is automatically reserved and purchased in real time from supercomputers and clusters of personal computers around the world. Another click, and data sets worldwide are scoured for all the Higgs two-photon events recorded so far. The interface invisibly converts all the datasets into a compatible format. One keystroke, and a menu pops up offering a suite of advanced visualization techniques that will allow the data to be analysed interactively with colleagues elsewhere.

Turning this dream into reality was the aim of a meeting at CERN last week between officials from the major European research councils, the European Space Agency and the European Bioinformatics Institute. The proposal that emerged would build on middleware developed from the existing Globus distributed software effort, a joint project of the Argonne National Laboratory in Illinois and the University of Southern California's Information Sciences Institute. The Linux operating system would be used throughout, says Williams, to ensure that software remains open-source.

By coordinating the national efforts that are being planned — Britain is likely to approve £100 million (US\$158 million) in funding for grids later this year — the proposal could quickly be scaled up to include other disciplines and industry.

"It is clear that there are people in other sciences very interested in doing the grids together," says Chris Jones, CERN's head of technology transfer.

Support for the development of grids is also awakening within industry. Jones last week discussed the grid proposal with a delegation of British industrialists, and Williams is optimistic that concrete industrial support for the proposal will be in place before the 10 May deadline.

Declan Butler

Uproar at Oxford as suspended professor returns to work

London

Two staff members have begun grievance procedures against the University of Oxford over its handling of their complaints against Roy Anderson, the Linacre professor of zoology.

They say that Anderson's apologies to staff who had accused him of offensive behaviour are "unsatisfactory". They have protested at the university's decision last week to let him resume teaching and research.

Anderson remains suspended as director of the Wellcome Trust Centre for Epidemiology of Infectious Diseases, as investigations continue into alleged business irregularities (see *Nature* 403, 695; 2000). Solicitors for zoologists Sunetra Gupta and Karen Day, whose complaints initiated the disciplinary proceedings, say that Anderson's apology has not removed a "sexual slur" against Gupta. It is alleged that Anderson had claimed she had an improper relationship with the head of the university's department of zoology, Paul Harvey.

The solicitors say that a statement from the university made "no reference

to the fact that complaints had been upheld and that Professor Anderson has been disciplined as a result".

The university says that written apologies have been sent to all concerned, but Harvey says that he has not received

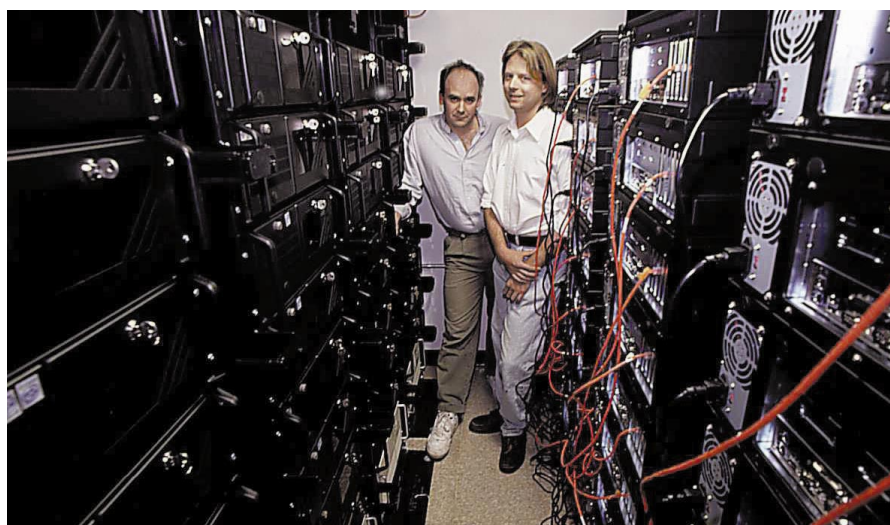


Gupta: apology "inadequate".

one. "I want a retraction [of the sexual slur], which I have not yet had," he adds.

News of Anderson's reinstatement was met by uproar at a meeting of senior staff which passed a unanimous vote of no confidence in him. Twenty-five zoology faculty members have sent a letter of protest to the vice-chancellor over the reinstatement of Anderson without their acceptance. Some staff members say he should be banned from appointment committees and that he should not be reinstated before financial and management audits are completed.

Anderson's lawyers said this week: "Disciplinary matters have come to an end, the university has issued a statement and Professor Anderson has signed up to that statement." **Natasha Loder**



Parallel evolution: Wheeler (left) and Janies survey their handmade, 129-box phylogeny-cruncher.

Homegrown computer roots out phylogenetic networks

London

If Ward Wheeler and Dan Janies suffer any computer glitches, they won't need to rely on the IT helpline. These two systematic biologists, based at the American Museum of Natural History (AMNH) in New York, should know exactly how to fix their powerful parallel computing system — they built it themselves from scratch.

Their machine is a testament to what can be achieved with some imagination, a 'can do' attitude, and a truckload of parts from your local Radio Shack. And it is already being used to slash the time needed to find the most likely phylogeny describing the evolutionary relationships between groups of organisms, or taxa.

Systematists like Wheeler and Janies view phylogenies as 'unrooted' networks. 'Rooting' the network to a putative common ancestor, producing the familiar textbook evolutionary tree, comes later.

Identifying the most plausible unrooted network can present a formidable computational challenge. There is just one unrooted network that expresses the relationship among three taxa. But after that, the numbers go up rapidly: with four taxa, there are three possible solutions; with five taxa, there are 15; with 10 taxa, there are 2,027,025, and with 20 there are around 2.21×10^{20} .

Rather than test every possible permutation, systematists use algorithms that select the ones that look most likely and ignore the majority — a compromise, perhaps, but one that produces solutions in less than a human lifetime. But even so, you need a powerful computer to crack difficult phylogenetic problems in hours, rather than years.

Wheeler and Janies have given themselves

a state-of-the-art computer for a fraction of the usual price. The secret is a massively parallel design, using off-the-shelf components, a freeware operating system, and homegrown phylogenetic software. The hardware is a stack of 129 boxes each fitted with a hard drive, a pair of 500-MHz Pentium III processors and 512 Mb of RAM, connected over a 100-Mbit switched network. One machine acts as the master that controls all the others.

"We made the thing from bare bones," says Wheeler. When they placed the order for the parts, he says, the incredulous electronics supplier thought that Wheeler and Janies ran their own computer shop. Bolting the bits together took a week and a half.

Fitting out a suitable room with air conditioning and switching cost US\$300,000; the hardware came to another US\$300,000. The operating system was free — the machine uses Linux, a freeware system similar to Unix. Funds came from NASA, the AMNH and the City of New York.

Wheeler is currently using the system to crunch through a 130-taxon insect database that combines data on morphology and molecular biology. There are 10^{154} possible networks for a data set this size. But the new system allows unprecedentedly fast and aggressive search strategies. "We found better solutions in hours than we had found in months of searching before," says Wheeler.

The system can be accessed remotely, and is being used by an increasing roster of collaborators working on a variety of problems. The homegrown parallel-processing software used by Wheeler and his colleagues can be downloaded from the Internet, at ftp.amnh.org/pub/molecular/poy and at ftp.amnh.org/pub/molecular.malign. **Henry Gee**

Irish researchers divided over how to spend huge new fund

Munich

In a remarkable reversal of fortune, Ireland's spending on basic research is set to increase by more than an order of magnitude. A new foundation is to be created to administer the money. But scientists are divided over plans for how the foundation will spend the windfall.

Around half of a new £1.95 billion (US\$2.5 billion) fund for research, technology and innovation in Ireland's six-year National Development Plan should feed directly into basic research. Researchers in biotechnology and information technologies will gain most, receiving £560 million.

A further £550 million is expected to be allocated by the Higher Education Authority to support research in universities across all disciplines. The goal is to create a 'knowledge base' to support the high-tech companies that have recently settled in Ireland, and to

lure the country's scientific diaspora home.

Science minister Noel Treacy boasts that the initiative will bring a "sea change of attitude" in a country that has long languished near the foot of Europe's league table for spending on research.

The foundation will fund biotechnology and information technology projects using international peer review. More controversially, Treacy says it will also "have the option to establish its own laboratories if necessary, to secure world-class performance".

Many scientists are deeply unhappy with the idea of investing in bricks and mortar. They point out that Ireland has recently invested heavily in its university infrastructure from a special £210 million fund that was set up last year.

"Monolithic, stand-alone institutes very easily turn into white elephants," warns Mark Keane, professor of computer science



MICHAEL ST. MAUR/CORBISO

Academic question: should new money go to established centres such as Trinity College?

at University College Dublin. "In the past, there has been a huge amount of unsupported talent in Irish universities: now is the time to support people, not places."

Many believe the money should be used for project grants and to create prestigious positions to entice the best researchers to Ireland. There is also some disquiet about the fact that the government has already allocated £50 million of the foundation's funds to support the Massachusetts Institute of Technology's Media Lab Europe, recently established in Dublin.

"When a government commits to increase research spending by one to two orders of magnitude then of course we should be celebrating and not arguing too much about the details," says David McConnell, director of the Smurfit Institute of Genetics at Trinity College Dublin. Only two years ago, Irish researchers were facing a situation in which their budget for projects could be wiped out at a moment's notice (see *Nature* 392, 529; 1998). But McConnell too, believes that the greater part of the new money would best be deployed in Ireland's existing universities.

The government likes the idea of independent institutes because they would give the initiative highly focused visibility. Frank Gannon, head of the European Molecular Biology Organization and formerly a professor at the University of Galway, says that a new, well-funded institute would be "very likely to develop a high-quality research ethos in Ireland".

Gannon is a member of the group that will help decide the new foundation's structure. Other members of this group are less enthusiastic about creating new institutes. They need to resolve their differences quickly, however. For accounting reasons, £25 million of the foundation's funds must be spent before the end of the year. **Alison Abbott**

Protests force primate farm to close

London

Britain's only supplier of primates for research has been forced to shut down after a campaign by animal-rights activists.

Primates are not bred for research in Britain. Shamrock Farm in West Sussex imported them from breeders in countries such as Mauritius and the Philippines, then sent them to laboratories. Once its remaining animals have been distributed to scientists, the facility will close.

In recent months, Shamrock Farm has seen confrontations between masked

protesters and police. The farm's managing director has suffered a series of attacks; most recently someone broke into her garage and set fire to her cars.

A total of 3,098 primates — most of them macaques — were used for research in Britain in 1998, the last year for which figures are available. More than half came directly from suppliers overseas. The closure of Shamrock Farm means that all primates used by British labs will in future be supplied in this way.

Shamrock Farm is the latest in a line of British suppliers of research animals forced to close after being targeted by animal rights activists. In 1997, Consort Kennels of Ross-on-Wye, a breeder of beagles, closed after a concerted campaign during which it was besieged by thousands of activists, its perimeter fence was breached and directors' homes were attacked. Last year, Hillgrove Farm in Witney near Oxford closed after letter bombs were sent to the farm's owner, Christopher Brown, and two employees.

"It's very clear that the activists will now go and target other animal breeders," says Mark Matfield, director of the Research Defence Society. "It's a worrying trend."

Since the Home Office regulates animal experimentation and oversees the Crown Prosecution Service and the police, says Matfield, its ministers could easily order a crackdown on activists. The Home Office was not available for comment. **Peter Aldhous**



Unlucky for some: the farm is to close after a series of anti-vivisection protests.

TIM OCKENDEN/PA

NEON to shed light on environment research

San Diego

Ten sites across the United States are to be linked in a network that will provide researchers with a continuous ecological health check. The National Science Foundation (NSF) is in the final stages of planning a \$100 million National Ecological Observatory Network (NEON), which aims to provide a computerized instrumentation infrastructure for decades of environmental research.

Last week, at the University of California at San Diego, ecologists, biologists, computer scientists and systematists met to refine plans for NEON. They see the network as a system that would provide an unprecedented real-time picture of ecological systems, recording data on biodiversity and climate and integrating this with data from other studies, including genomics, soil chemistry and hydrology.

"This is a once-in-a-career opportunity to create an infrastructure for ecology," says Jim Beach, assistant director for informatics at the University of Kansas' Natural History Museum. Late last year, the National Science Board, which governs the NSF, approved the NEON concept in the agency's budget for long-term infrastructure projects. This is the first time the NSF's biological division has had such a major equipment budget request; previously, such largesse has been the exclusive domain of disciplines like physics and astronomy.

After the first NEON workshop in January in Florida, a group of 26 academic

researchers and four NSF officials issued a report, saying: "The goal of NEON is to develop distributed infrastructure for environmental biology, ranging from molecular genetics to landscape-level study. The infrastructure should advance the study of continental- and regional-scale issues."

Last week, more than 30 meeting participants discussed the equipment needed to make NEON a successful scientific system. Field biologists, for instance, want inexpensive wireless systems for data transmission from field sites to a central network, digital field guides on species and remote access to centralized instruments — including microarrays for genomic studies, mass spectrometers and high-performance computers.

Each observatory will cost around \$10 million to create, and NSF officials say they will provide \$1 million a year in operational funds for each facility. The agency also



Field work: Konza Prairie, one of the LTER sites that could be included in NEON.

expects to have a pool of research funds — the amount has yet to be determined — for proposals for research through NEON. Once in place, the observatories should operate for 30 years.

NEON funding is now included in the Clinton Administration's budget proposal for the 2001 fiscal year that begins in October. Congress still has to approve the funding, and building the entire network will require a spending package spread over 6 years.

In the initial round of the competitive process for NEON designation, later this year, NSF officials aim to fund three observatories. The next round, probably in 2001, will designate four more observatories, and the final three observatories will be picked sometime after that. Bids are expected to come from consortia of universities, museums and research institutes.

At the workshop last week, participants urged the NSF to include other ecological facilities in the consortia — such as the 24 Long-Term Ecological Research (LTER) facilities already supported by the NSF, plus survey sites run by the US Geological Survey and the US Forest Service.

The birth of NEON comes after more than a year of prior planning and workshops on a related NSF concept called the Biodiversity Observatory Network. This aimed to monitor a range of sites in each major region of the United States, but lacked the NEON's emphasis on networking. Last year, however, NSF officials decided that NEON should proceed first.

Rex Dalton

ALAN K. KNAPP

Are AIDS dissidents advising South Africa?

Cape Town

Leading AIDS dissident David Rasnick has claimed that Thabo Mbeki, the president of South Africa, requested his scientific opinion on eight questions related to HIV and AIDS last January.

Rasnick is one of a small band of scientists who claim that HIV does not cause AIDS. Mainstream AIDS researchers are deeply concerned at the suggestion that he is advising the South African government.

Mbeki has also allegedly put the same questions to health minister Manto Tshabalala-Msimang, who has caused concern by refusing to clarify whether she accepts that HIV is the cause of AIDS.

Rasnick has posted the text of what he claims are Mbeki's questions on the 'Virusmyth' website (www.virusmyth.com), along with his reply, co-written with another AIDS dissident, the historian Charles Gesheker of California State University.

According to Rasnick, Mbeki wants to provide a public forum where proponents and critics of the HIV hypothesis can present the evidence for and against the assertions that AIDS is contagious and sexually transmitted, that HIV causes AIDS, and that anti-HIV drugs help people with the disease.

Rasnick further claims that Mbeki is asking other world leaders, including US president Bill Clinton, UK prime minister Tony Blair, and German chancellor Gerhard Schroeder, to join him in a discourse about these issues. Rasnick's Internet posting coincided with an announcement by Tshabalala-Msimang on the establishment of a panel of experts to investigate various aspects of AIDS (see *Nature* 404, 115; 2000).

With Peter Duesberg, his colleague at the University of California, Berkeley, Rasnick co-authored an article entitled "The AIDS dilemma: drug diseases blamed on a passenger virus", published in the journal

Genetica in 1998. Duesberg has confirmed that he has not been asked to sit on Tshabalala-Msimang's panel, but says he has written to Mbeki confirming his stand on the HIV-AIDS hypothesis. He indicated to *Nature* that Rasnick had been approached to participate on the panel.

Tshabalala-Msimang is coming under increasing pressure for her unorthodox views on AIDS. She was booed by angry activists at a dinner in Durban last week, at which, referring to the use of antiretroviral drugs, she said that she didn't want to "plunge into something I don't understand".

This followed a stinging attack by the acting Constitutional Court judge Edwin Cameron, who is HIV-positive. Cameron criticized the government's policies on AIDS as causing "considerable grief and confusion", and also questioned Tshabalala-Msimang's competence as a minister.

Michael Cherry

German research agency stifles creativity

Munich

Researchers such as Mark Benecke should be the future of German science. With most of his contemporaries still only just finishing their PhDs, this 29-year-old entomologist at the University of Cologne is already applying for research grants under his own steam. Yet earlier this year, his application to the Deutsche Forschungsgemeinschaft (DFG), the country's main grant-giving agency for university research, was turned down after an eight-month wait. "The DFG had huge difficulties in assigning my proposal to appropriate referees," he says.

Benecke's field is forensic entomology, which can, for example, reveal how long a corpse has lain undiscovered by studying the insects on it. And it is typical of the novel research areas that, according to many scientists, the DFG seems unable to assess. They charge that the agency's aversion to taking risks, lengthy review procedures — in some cases more than twelve months — and inability to deal with interdisciplinary proposals threaten career opportunities for young researchers.

Ultimately, some researchers warn, the agency's outmoded procedures could weaken Germany's position in cutting-edge areas of science. And feelings are running so high that the debate has spilled over into the science and correspondence pages of leading German newspapers. Germany risks the "migration of the best minds of its next generation of scientists", says Ralf Müller, a German electrical engineer at the University of Princeton in New Jersey.

Many question the DFG's ability to handle interdisciplinary research applications. Its elected external referees are organized into narrow, discipline-orientated boards, and critics claim they fail to keep pace with fast-moving areas such as biomedicine, ecology, informatics or materials research. Time is lost because the DFG has trouble finding referees for innovative or cross-disciplinary applications, they say.

"The key problem is that funding decisions are made by a small group of over-worked honorary referees, nominated by the traditional German scientific societies," says Michael Cross, a young British molecular biologist at the University of Leipzig.

The DFG's review system "is not as flexible as it should be", agrees Gerhard Neuweiler, a professor of zoology at the University of Munich and former president of Germany's science council, the Wissenschaftsrat. "Unfortunately, this tends particularly to disadvantage innovative researchers in the most competitive areas."

Peter Uetz, for example, applied unsuccessfully to the DFG in 1997 for a project in functional genomics. Uetz then turned to the



Benchmark of despair: innovative German researchers find it hard to get grants.

German Academic Exchange Service, and eventually received a grant allowing him to do research at the University of Washington in Seattle. His first results were published last month in *Nature* (403, 623; 2000).

"Ironically, the DFG turned down my proposal at a time when its president, Ernst-Ludwig Winnacker, was strongly emphasizing the strategic importance of functional genomics in Germany," says Uetz.

Many German scientists look enviously across the Atlantic, where they feel funding agencies are prepared to support risky projects, trusting that one of them will turn out to be a big breakthrough. They criticize the DFG's tendency to cover itself against possible failures at the expense of missing out on the new and unexpected. Faced with such conservatism, young researchers such as Müller and Uetz are voting with their feet.

DFG officials defend their record, pointing out that decisions on grant applications take on average between five and six months

—comparable to organizations such as the US National Science Foundation and the British Medical Research Council. But critics say this respectable figure includes grants in the humanities and traditional scientific disciplines, where the DFG's performance is not under attack.

The DFG had promised organizational changes after similar criticisms were levied by an external evaluation last year (see *Nature* 399, 395; 1999). But many scientists are disappointed with the lack of progress, and some suggest that a more pluralistic system may be the answer. "Flexibility would increase very quickly if the DFG lost its monopoly," says Oliver Kempfski, a professor of medicine at the University of Mainz.

Unlike its counterparts in the United States and Britain — and private foundations such as the German Volkswagen Foundation — the DFG has no established mechanisms to 'fast-track' applications in hot research areas such as computer-based functional genomics or molecular medicine. "In highly competitive areas, one week can be a very decisive factor," says Neuweiler.

Cross suggests that the DFG should allow interdisciplinary projects to be reviewed simultaneously by referee boards representing different disciplines. In particularly complicated or urgent cases, he adds, applicants should be given the chance to attend referees' meetings to explain their ideas face to face.

Eva-Maria Streier, the DFG's spokeswoman, says the agency is open-minded about this idea. She points out that the agency is already experimenting with such parallel reviewing, although merely in addition to its normal procedures.

Unless the DFG makes urgent reforms, say some researchers, Germany risks losing international competitiveness. "Eventually you will find yourself in the second division," says Cross.

Quirin Schiermeier and Patrick Weydt

EU centres saved from 'catastrophe'

London

European research ministers intend to reverse controversial rules which have prevented the European Union's Fifth Framework Programme of Research (FP5) from supporting the running costs of European research facilities.

Life-science research facilities were thrown into crisis last year when officials at the European Commission unexpectedly said that institutions such as the European Bioinformatics Institute (EBI) in Cambridge and the European Mouse Mutant Archive (EMMA) at Monterotondo near Rome could not be funded as service facilities.

Both facilities, which life scientists hold to be essential for the development of their disciplines, had been set up with FP4 support, with the understanding that funding would be continued in future framework programmes (*Nature* 402, 3; 1999).

The Council of Ministers had authorized support for infrastructures for a wide range of activities in FP5. However, commission officials interpreted the text of the FP5 document as excluding running costs for research facility infrastructures.

But José Mariano Gago, the Portuguese science minister, told an informal meeting

in Lisbon last week, where European Union (EU) research ministers met with heads of European science organizations and a handful of Nobel laureates, that he is not sure the interpretation was legal.

The ministers will now ask the commission to reverse its decision not to distribute money to the service facilities, said Gago. If the commission disagrees, he said, then the ministers will take a decision at their next formal meeting in June to unambiguously change the rules.

"Clearly the decision was not what the council of ministers intended," Gago told *Nature*. "All the ministers think the situation is quite unfair and should be resolved immediately. The ministers did not decide [to stop funding for EBI and EMMA]... it was a decision by the commission which was taken in a bureaucratic way during the summer holidays."

Gago says the ministers believe infrastructure support to be "essential" for research and development in Europe.

Glauco Tocchini-Valentini, secretary general of the European Molecular Biology Conference, welcomes Gago's statement. "Facilities like EBI and EMMA are close to catastrophe, so intervention at the political level is urgently needed."

But the commission remains reluctant to "create a precedent for direct financing of infrastructures", says a spokeswoman. The FP5 budget is limited, and demand, particularly in the life sciences, is growing, she says. The commission is setting up a working group to look into other options for saving EMMA and EBI.

The Lisbon meeting was the first time that EU research ministers had met with such a wide range of representatives from the scientific community. The EU no longer has a formal European scientific advisory structure. The advisory European Science and Technology Assembly, which was associated with FP4, was dissolved two years ago by former research commissioner Edith Cresson (see *Nature* 394, 817; 1998) and has not been replaced.

Nobel laureates at the meeting heavily criticized EU peer review systems and called for transparency and efficiency. Gago told *Nature* he thought a formal mechanism was needed for closer cooperation between the scientific community, the council and the commission. "The European research organizations and the national research councils must be part of the normal consultation process of the council and the commission," says Gago. **Natasha Loder**

New form of hydrogen power provokes scepticism

Washington

A company claiming to have made a revolutionary breakthrough in chemistry and energy production by creating a novel form of hydrogen has threatened several prominent physicists with possible legal action unless they stop disparaging the science behind the claim.

A law firm representing the energy company BlackLight Power, Inc. of Cranbury, New Jersey, sent letters earlier this month to Nobel laureate Philip Anderson of Princeton University, Michio Kaku of the City University of New York, Paul Grant of the non-profit energy agency EPRI and Robert Park of the American Physical Society, requesting that they stop making defamatory comments in the press about the company and its president, Randell Mills.

BlackLight has already attracted more than \$20 million in private investment to back its proprietary chemical process. According to Mills, this process has generated energy far in excess of that put into the system. Underlying the process is Mills's theory that hydrogen atoms can be made to exist below their ground state in a form he calls "hydrinos".

The four scientists cited by BlackLight have been quoted in the *The Village Voice*, *Dow Jones Newswire* and other publications as dismissing the claim because it violates established principles of physics. Kaku, commenting on the company's investors, which include several large utility companies, was quoted by the *Dow Jones Newswire* as saying "There's a sucker born every minute".

The negative publicity comes at a bad time for BlackLight, as the company is considering a public stock offering this year. Mills accuses his critics of "trying to destroy our business", and bristles at the charge that he has produced no data to back his claims.



He points to a conference presentation he gave last year at a regional meeting of the American Chemical Society, and says he intends to present at the society's national meeting this month.

He also says he is preparing papers for submission to major scientific journals, and that others have replicated his results and are also submitting to journals.

So far, though, Blacklight's results have been published only on the company's website (www.blacklightpower.com) or in journals that many mainstream scientists say lack rigor and are dominated by other researchers investigating unconventional — some say impractical — forms of energy.

Nor has the company's recent award of a US patent for "Lower-energy hydrogen methods and structures" impressed the critics. Grant, an expert in high-temperature superconductors, was quoted by Dow Jones as saying, "A patent means nothing. It carries no weight as scientific validation." The patent examiners based their decision on presentations by BlackLight, according to Mills.

Park says, "the issue is not whether their stuff is out there for review. The issue is whether anybody believes it, and whether people who don't believe it have a right to say they don't believe it." He continues to discount BlackLight's claims as "pure boloney", and will say so in his book, *Voodoo Science*, due to be published by Oxford University Press this spring. "There'll be no changes," he says.

Despite the implied threat in his letter, BlackLight's attorney Michael O'Hayre says that "we're not interested in stifling any free and open debate. Right now we're just investigating what to do."

Park says he has turned the letter over to the solicitor at the American Physical Society, and is confident that, should BlackLight decide to sue, the courts would side with the physicists. The scientific community would also be likely to rally round the defendants, he says, just as they did a decade ago when lawyers for proponents of cold fusion sent out threatening letters.

Although legal threats in scientific disputes are surprisingly rare, Park says scientists can be "pretty easy to intimidate". With legal fees to defend against a libel charge sometimes running to tens of thousands of dollars, he admits some scientists could decide that it's not worth the risk speaking to the press about controversial research. And that, he says, would "leave the public vulnerable". **Tony Reichhardt**

Art imitating high-energy physics

Munich

Some of Europe's leading conceptual artists are taking part in an extraordinary cultural experiment at CERN, the European Laboratory for Particle Physics.

The artists, who include Turner Prize winners Anish Kapoor and Richard Deacon, have been brought to CERN to learn about high-energy physics and will respond by creating an original piece of art during this year. These artworks will be exhibited at galleries across Europe next year in a roadshow called "Signatures of the Invisible" — a reference to the fact that subatomic events can only be described by their mathematical signatures.

So far, most of the artists seem to have been more impressed by CERN's physical environment, its engineering tools and exotic materials, than by the details of theoretic physics. After visiting CERN, Deacon said: "We need to listen to each other, but not necessarily to understand. The misunderstanding in both directions can be creative."

British artist Patrick Hughes is known for his work on 'reverse perspective'. His three-dimensional scenes appear to shift dramatically in perspective as observers move around them. Hughes says he may build a large three-dimensional model depicting the huge doors to CERN's subterranean experimental chambers, opening onto a view of the mountains that surround Geneva. Deacon, a sculptor, says he is toying with ways of



A new perspective: this work, *Hoppera* (1999), is by Patrick Hughes, who is part of the CERN project.

exploring "the edges of things". He is impressed with CERN's exotic materials "not just for their properties, but also for their metaphoric significance".

However, some of the artists are getting to grips with physics. Bartolomeu dos Santos, a Portuguese artist, has produced many pieces of public art by etching the surface of materials. He is using CERN's electron welding facility, normally used to produce distortion-free welds for high-precision instruments, to etch representations of event

displays and theoretical formulae into steel.

Film-maker Ken McMullen, who is organizing the project, will be addressing another phenomenon of physics: crumpling. His film will record the way that metals exposed to the same pressure under the same conditions always crumple in a different way. "The only variable is time," says McMullen.

CERN director-general Luciano Maiani admits to being bemused by the project, but says he is "happy to let it carry on around CERN's periphery".

Alison Abbott

Older women scientists fight USGS over layoffs

San Diego

Women scientists who lost their jobs at the US Geological Survey (USGS) office in Denver, Colorado, during a major layoff have filed a discrimination complaint with federal authorities. They allege that they were targeted for discharge because of their age and gender.

This is the latest in a series of legal fights that have continued since the 1995 'reduction in force', when 525 of the 1,850 scientists in the USGS geology division, which is spread over different sites, were laid off, retired or diverted to other jobs. Since then, more scientists have left.

The complaint was filed with the Equal Employment Opportunity Commission (EEOC) in January by two former scientists on behalf of up to 20 women.

The two women — Constance Throckmorton and Kathryn Nichols — claim that female geologists older than 40 were "affected at a rate nearly twice that of male counterparts". Of the 44 women scientists in the Denver division, 45 per cent

were laid off, while only 23 per cent of the 215 men lost their jobs, they say.

Nichols, who had worked at the USGS for nineteen years, was a year from retirement eligibility when she was laid off. She says that her history of challenging "an old-boy network" had made her unpopular with top USGS officials who "didn't want anyone challenging old [scientific] conclusions". Nichols, who received her doctorate at Stanford University, says: "I was an intellectual threat."

In support of their belief that uneconomical overstaffing was not the genuine reason for the redundancies, the women cite an e-mail which USGS Chief Geologist, Patrick Leahy sent to all geological staff in early January, in which he says that the "top concern" of the USGS is to recruit new scientific staff, which the agency is doing by hiring a "modest" number of postdoctoral fellows.

Thomas Fouch, the regional geologist in Denver, says that despite the sharp blow to morale caused by the layoffs, the USGS is

"in good [scientific] shape", in contrast to allegations by some dissatisfied staff. The number of scientific publications published by the USGS has not decreased since 1995, he says.

The move was made necessary by budget cuts, he claims. He adds that the workforce reductions have allowed the USGS to implement modern employment standards because future staff will be hired on short-term contracts. The laid-off USGS scientists thought they had "a job for life [as] an entitlement", says Fouch.

USGS director Charles Groat and other top officials at the agency's headquarters in Virginia declined to comment on the discrimination claim. The EEOC will review the allegation, conduct a hearing and issue a ruling. The women can take their complaint to the federal court if they are dissatisfied with the ruling.

Meanwhile, one long-time geologist in Denver privately describes the current situation as follows: "The mood is very pessimistic; morale is low."

Rex Dalton

NASA gets green light for Triana satellite to photograph Earth

Washington The US space agency NASA will resume work on the Triana satellite project after a panel of the US National Academy of Sciences concluded that Vice-President Al Gore's controversial proposal to launch a satellite to photograph the Earth continuously has scientific merit.

Congress last year ordered NASA to suspend work on the project pending the result of the review. The White House science office was so eager to trumpet the good news that presidential science adviser Neal Lane issued a statement that "Triana... should not be kept earthbound" several hours before the academy's press office posted the report on the web.

The space agency is still prohibited by Congress from launching before January 1, however — after the presidential election.

US cancer body plans pathogenesis database

Washington The National Cancer Institute is to compile a set of small-molecule probes to explore cancer pathogenesis. By pitting the probes at different targets, then recording

binding data in a public database, Richard Klausner, NCI director, hopes to build a "chemical toolbox" for cancer information, akin to GenBank, which holds genomic information. The project will help identify the most effective drugs against different forms of cancer. The NCI will spend \$15 million to begin the endeavour.

Harvard supports Lebanese genetics lab

London Lebanon has opened a new Genetics Research Laboratory based in the Chronic Care Centre in Baabda. The laboratory has been established to address congenital and genetically determined chronic disorders.

The laboratory is the result of close collaboration with the Harvard School of Public Health (HSPH), which will allow work on environmental and genetic epidemiological studies to look at the relative contribution of genetic and environmental factors to complex diseases.

Pierre Zalloua, formerly a research associate at the HSPH, is to head the lab, working with Xiping Xu, an environmental health epidemiologist at the HSPH.

The laboratory hopes to promote and provide research and development in molecular genetics in the Lebanon and elsewhere.

BSE claims its youngest victim so far

London New-variant Creutzfeldt-Jakob disease (vCJD), the fatal condition linked to the consumption of BSE-infected beef, has claimed a 15-year-old victim. Claire McVey, from Ilfracombe, Devon, UK, died in January. She is included in the latest set of monthly confirmed and presumed cases of vCJD.

Peter Smith, an epidemiologist at the London School of Hygiene and Tropical Medicine, says that cases could emerge among anyone in Britain who ate products such as hamburgers in the late 1980s — and even the early 1990s, given that controls to exclude infected tissues from the food chain may not have been adequately enforced.

Eleven cases of vCJD have now been reported for 1999, six less than the figure recorded for 1998. However, epidemiologists say it is still too early to make any firm predictions about the eventual human toll, which so far stands at 52.

Italian universities open to review

Genoa Some of Italy's more outward-looking universities opened themselves up to external assessment for the first time last week. A committee appointed by the Association of

European Universities (CRE) has visited three campuses, in Genoa, Rome and Camerino, to assess their management systems. A final report will be presented to the rector of each university and made available to the public in autumn.

CRE's function is purely consultative. But university rectors consider the evaluation to be a useful means of encouraging reform of the system of allocating resources for teaching and research.

Inventor of 'herbal fuel' arrested for fraud and theft

Madras A self-taught scientist who became famous in India three years ago as the inventor of a herbal fuel for cars has been arrested. Federal investigators in Madras allege that Ramar Pillai cheated the public by selling stolen industrial chemicals as motor fuel.

Pillai, who manufactures his fuel in Madras and has 11 retail outlets, has been accused by the Central Bureau of Investigation of fraud, theft and sale of fuel comprising a mixture of chemicals potentially damaging to vehicles. Investigators seized more than 10,000 litres of Ramar Herbal Fuel, containers of chemicals, additives and a large sum of cash.

The spokesman at the Central Bureau of Investigation said that the so-called herbal

fuel was nothing more than a mixture of benzene and toluene, industrial organic chemicals. He said that Pillai and some officials in a public sector firm were systematically stealing the chemicals from refineries and making a hefty profit. Pillai's product was only a little cheaper than ordinary fuel, but he was still reported to be doing good business.

Rutherford-Appleton wins synchrotron race

London The new Anglo-French synchrotron source will be based at the Rutherford Appleton Laboratory near Didcot, Oxford.

The announcement brings to a close months of speculation about the siting of the new facility (*Nature* 402, 111 & 451; 1999). It will be a severe disappointment to those campaigning to keep the next-generation source of X-ray light at the home of the existing national source, at Daresbury Laboratories near Manchester. Staff, union officials and several members of parliament from the north of England have been campaigning to win the new source for nearly six months.

The UK government said the decision had been "difficult". One of the sweeteners offered is a £5 million (US\$7.9 million) upgrade of the existing facility at Daresbury.

Four new beamlines will extend Daresbury's lifespan from five to seven years. This means that the existing synchrotron source will operate for a while in parallel with the new synchrotron.

Switzerland rejects ban on human *in vitro* fertilization

Munich The Swiss population turned down a proposal to ban *in vitro* fertilization of humans in a plebiscite held last Sunday by a majority of 75 per cent.

The proposal was submitted by feminist and religious groups that have been campaigning against reproductive medicine since the birth of the first *in vitro* fertilized baby in Switzerland in 1985, although the number of such groups has decreased. On average, two babies conceived by methods of reproductive medicine are born each day in Switzerland.

Correction

'Celera in talks to launch private sector human proteome project' (*Nature* 403, 815–816; 2000).

The former Perkin-Elmer Corporation has changed its name to PE Corporation, and is the parent company of PE Biosystems and Celera Genomics. The brand name Perkin-Elmer now belongs to EG&G Corporation, which acquired The Perkin-Elmer Corporation's analytical instruments business in May 1999.

There's enough food for everyone, but the poor can't afford to buy it

Sir— The existence of malnourished and hungry people has been used repeatedly in this journal and elsewhere as a justification for biotechnology and for the production of more food^{1,2}. This assumption supports a main policy plank of the Rockefeller Foundation food biotechnology programme² and other major international and charitable institutions.

Yet there are good reasons to be sceptical of the equation “more food equals less hunger”.

The world produces more than enough food at present to feed everyone, but nevertheless many people still starve or are malnourished^{1–3}. As economist and Nobel laureate Amartya Sen has pointed out, it is poverty, not a physical shortage of food, that is the primary cause of hunger in the modern world⁴.

The political and economic reasons don't change: the amount of food that Ireland, for example, exported to Britain during the potato famine of 1845–46 would have been sufficient to feed those who starved. The root cause of the 1974 Bangladesh famine was a flood that displaced people from their jobs; more food was produced that year in Bangladesh than in surrounding years, yet — unable to earn money to buy it — up to 1.5 million people starved to death⁴.

Partial solutions such as local production of food, as suggested by Conway and Toenniessen², cannot circumvent economic reality. Even the World Bank has concluded that the problem of hunger can only be solved by “redistributing purchasing power” to the hungry⁵.

What about the state of food supplies in, say, 2040, when it is predicted that there will be ten billion people compared with today's six billion? In absolute terms, the world already produces enough food to feed ten billion people — it's just that most of it is fed to animals (this accounts for 80% of all arable crops in the United States, a figure close to the world average⁶). If the area of arable land devoted to crops for human consumption were doubled to 40%, this need not drastically affect supplies of meat or dairy produce, since farm animals also eat other, non-crop foods such as grass in the summer, silage and hay in the winter.

Clearly, what is missing is the “purchasing power” of the poor.

Jonathan R. Latham

25 Barrack Road, Exeter, Devon EX2 5ED, UK

1. Trewavas, A. *Nature* **402**, 231–232 (1999).

2. Conway, G. & Toenniessen, G. *Nature* **402**, Supplement C55–C58 (1999).

3. Simms, A. *Selling Suicide: Farming, False Promises and Genetic Engineering in Developing Countries* (Christian Aid, London, 1999).
4. Sen, A. K. *Poverty and Famines: An Essay on Entitlement and Deprivation* (Oxford Univ. Press, Oxford, 1983).
5. World Bank. *Food Security for the World* (statement prepared for the World Food Summit, 1996).
6. *Agriculture Statistical Yearbook* (Eurostat, Luxembourg, 1996).

Spanish university study ignores research

Sir— As Xavier Bosch reports (*Nature* **402**, 848; 1999), the recently publicized ranking of Spanish universities by quality may not be the first such list, but it has had by far the most impact on Spain's mass media. The repercussions are likely to encourage the growth of an evaluation culture in Spanish universities, which is positive. However, I believe the importance of this study has been exaggerated and people have made judgements without evaluating the 71 indicators used by the authors.

I do not see what factors such as “age of the university” or “percentage of women among first-year students” can reveal about the quality of an institution (even though they favoured my university: it is 505 years old and more than two-thirds of its students are women). I did not find any indicator of research quality and there was no evaluation of scientific papers published in leading journals. Nor were patents, contracts or research projects considered.

Given that a university carries out both teaching and research, I estimate that this study has done only half the work.

Jorge Mira-Pérez

Department of Applied Physics, University of Santiago de Compostela, E-15706 Santiago de Compostela, Spain

Not too late to apologize

Sir— The editorial “Hollow apologies should be avoided” (*Nature* **403**, 813; 2000) gives a correct account of the efforts of the Max Planck Society (MPS) to explore the history of its predecessor, the Kaiser Wilhelm Society, during the Nazi period. It discusses MPS president Hubert Markl's statement that it is not within the moral authority of those who did not take part in Nazi experiments to apologize on behalf of those who committed these crimes. It suggests apologies may be due, rather, for the MPS's ignoring of the issue until quite recently.

The MPS can agree with this. However, President Markl has already publicly apologized. In 1998, at a ceremony for the fiftieth anniversary of the society's foundation, he said: “I consider it my duty to offer a public apology for that which the

Max Planck Society may have failed to do in the face of its responsibility for the consequences of its prehistory during the Third Reich — even were it only that the Society has done too little to explore this prehistory for too long.” In the same speech, Markl condemned the actions of German scientists against Jewish and other victims of the terrible Nazi past in more detail. The full text is on the Internet (http://www.mpg.de/jubilae_e.htm).

The question of public statements to survivors can only be decided after the independent commission of historians has given its advice to the president. But the Max Planck Society is grateful for this chance to let your readers know that it has acted — though maybe regrettably late — exactly as advised in your editorial.

Bernd Wirsing

Max Planck Society, Max-Planck-Gesellschaft, Postfach 10 10 62, 80084 München, Germany

DoE still involved in the Human Genome Project

Sir— I would like to correct the erroneous impression given by the News article “US energy agency pulls plug on role in genome project” (*Nature* **404**, 4; 2000). The Department of Energy (DoE) is not pulling the plug on its role in this important project. As part of the international human genome project, DoE is responsible for determining the DNA sequence of human chromosomes 5, 16 and 19.

DoE's commitment to completing this project has not changed: we will deliver, as promised, a working draft sequence of these three chromosomes in 2000 and their complete, high-quality sequence by 2003. We will continue to coordinate our human DNA sequencing efforts with our partners: the National Human Genome Research Institute at the National Institutes of Health and the Wellcome Trust in the UK.

Obtaining a complete, high-quality sequence of human DNA is only one of the goals of the Human Genome Project. The DoE biological and environmental research advisory committee (BERAC) has been asked to identify the next most important contributions it can make in other goals and is already investing in the development of the tools needed to exploit fully the value of knowing the complete DNA sequence of the human and other organisms.

DoE will continue its sequencing efforts in model organisms, including the mouse and various microbes, whose DNA sequence information contributes to DoE missions in bioremediation and carbon sequestration.

James F. Decker

Office of Science, Department of Energy, Washington, DC 20585, USA

Questions suspended in the ether

A recent discovery of early writings by Heinrich Hertz sheds light on his work.

Über die Constitution der Materie (On the Constitution of Matter)

by Heinrich Hertz

Springer: 1999. 171 pp. \$53, £34

David Hyder and Heinz Lübbig

Heinrich Rudolf Hertz (1857–94) is best remembered today for the unit of frequency named after him. But at the time of his death, at the age of 37, he had achieved international recognition for his work on electrodynamics, above all, for his experimental confirmation of James Clerk Maxwell's electromagnetic field theory.

In 1878, when Hertz began his scientific career in Berlin under Hermann von Helmholtz, three theories of electrodynamics were in play (roughly, those of Maxwell, Wilhelm Weber and Franz Neumann); these Helmholtz had reformulated into a single potential theory. As Hertz explained in his *Electric Waves* (1892), each theory was associated with a distinct 'picture' of its physical significance — the traditional concept of force as action at a distance; a mixed picture in which such forces were supplemented by the actions of an intervening medium; and Maxwell's 'pure' picture, in which action at a distance was replaced completely by a field. Hertz felt his achievement lay largely in distinguishing these pictures from each other and also from the mathematical theories to which they gave rise.

Hertz's intellectual development — when and why he became a convinced Maxwellian, and the connection between his electrodynamics and later work on mechanics — has proved hard to trace, in part because he died so young. The introductions to his two books — *Electric Waves* and the posthumous *Principles of Mechanics* (1894) — are among the few sources in which he speaks generally about his field and the significance of his work. To these we may now add a third, early work comprising a series of lectures delivered at a critical juncture in his career. Albrecht Fölsing, while researching his biography of Hertz (*Heinrich Hertz: eine Biographie*; Hofmann & Campe, 1997), discovered a manuscript entitled *Über die Constitution der Materie* in papers belonging to Hertz's family. This he has now edited and published with an introduction.

The material originates from the summer of 1884, when Hertz held a lecture series for non-specialists at the University of Kiel, where he had recently taken up a post as assistant professor. The course was sparsely attended, and in this informal setting he allowed himself to speculate on the future of field theory, to consider a wide variety of



Hertz: became a convinced Maxwellian.

possible experiments (one of which led to his famous wave-generation experiments), as well as to reflect on fundamental philosophical issues in physics. Hertz seems to have intended working the manuscript of the lectures into a book, but dropped the project and did not return to it. Even after his death, the editors of his papers contemplated publishing it, but despite enthusiastic support from Max Planck, those plans were also abandoned and the manuscript was forgotten.

The title of the work may strike the modern reader as misleading, for the first part seems not so much an investigation into the properties of matter as a long reflection on the nature of force. This concern is at the core of a formal paper on which Hertz was hard at work, and which he published later that year. In it, Hertz derived Maxwell's equations from those of 'conventional electrodynamics' by adding supplementary assumptions. He concluded by observing that, although the Faraday–Maxwell picture of the ether would lead directly to these equations, he had deliberately avoided departing from this, the 'simplest' interpretation of the equations.

Similarly, the discussion in the first part of the lectures is framed as a conflict between two fundamental concepts: the standard one, which views forces as instantaneous relations between arbitrarily distant masses; and what Hertz calls "the potential theory", to which he apparently subscribes. On this view, apparent actions at a distance result from local actions of a medium — from changes "in the state of space".

According to the potential theory's account of, say, gravitation, "the planets are attracted to the sun because they find themselves in a space which drives them toward

the sun by virtue of a peculiar state into which it has been put". This change in the state of space ("be it a disturbance, a tension or something else") propagates in time "from particle to particle". Explaining how this is possible is by definition an inquiry into the nature of matter, for "if there is a change in space, we cannot conceive of this otherwise than by there being something which is changed, and which we can call matter". In this sense, questions concerning the nature of force are also concerned with a special kind of matter, namely that of the ether.

Hertz was not the first to propose that all forces should be viewed as temporally propagated fields; nevertheless, the depth of his conviction in this early work is surprising. It certainly sheds light on his remark, in the introduction to *Electric Waves*, that "the philosophical, in a certain sense the most important result" of this later research was that it offered the first proof of "the propagation in time of a supposed action at a distance". It also suggests that the *Principles of Mechanics* — displacing observed forces onto the actions of hidden masses — fulfilled an early ambition: to explain the hidden material substrate of force fields by reducing them to local changes in the state of the ether.

The second section of the lectures deals with the essential properties of ponderable matter. The introductory discussion follows closely the pattern of Immanuel Kant's *Metaphysical Foundations of Natural Science* (1786), which Hertz had recently been reading; he seeks to determine what, if any, properties of matter are essential to our conception of it. Must matter be conceived as extended, as movable, as infinitely divisible, as compressible or incompressible? Must it be conceived as the source of a central force, as Kant had maintained? Must it possess inertial mass? Hertz follows Kant in claiming that the physical sciences presuppose a discipline which specifies how basic notions such as matter and force are defined and applied. He gives striking evidence for the need and usefulness of a true "science of the constitution of matter". The next chapter opens with questions concerning inertia and gravitational force, and Hertz argues that the tacit assumption that inertial and gravitational mass are equivalent is a conceptual error. The two notions are not equivalent, and since they are found to coincide empirically, this "wonderful enigma" cries out for an explanation.

In the rest of the second part, Hertz develops his question on the divisibility of matter by examining the state of contemporary atomic theory. The most interesting theme is Hertz's insistence that the very evidence in

favour of atoms must be interpreted as evidence for their inner complexity. The presence of absorption and plasma spectra alone suggests that atoms must contain vibrating elements, which we may think of as complexes of balls connected by rods and rubber bands, or as a minute planetary system, or as elastic solids.

Another problem presented by atomic theory concerns the possibility of extending the concepts of our everyday experience to matter on the atomic scale. Every attempt to do so is at best a vicious circle (for we cannot explain the macroscopic property of elasticity merely by reintroducing it at the microscopic level) and at worst a contradiction (for the sensible properties of matter are by definition insensible at the atomic scale).

Hertz responds to this problem with the earliest instance of his picture-theoretical concept of science: we have no choice but to use some sensory pictures in representing the subatomic world to ourselves, he says, and indeed they will involve superfluous, if not contradictory, characteristics; however, these pictures are only ancillary to our system of conceptually defined magnitudes, "which are connected among themselves and with the macroscopic qualities of matter by rigorously defined mathematical relations".

Unfortunately, Hertz's manuscript breaks off before he can conclude the section on ponderable matter. The first section also ends abruptly with a list of questions he did not get round to. Some of these bridge the topics of both parts: does the ether have inertia? Does it have mass? Is it composed of atoms? These questions are of pressing concern, because the idea of reducing forces to actions of the ether will at some point have to explain its mechanical properties. Indeed, one may suppose that the *Principles of Mechanics* was intended to lay the theoretical foundations for a deeper investigation into these questions. Confronted with the problem of explaining how we might conceive of the "changes in the state of space" required to eliminate forces, Hertz postulated the existence of hidden masses that were, in his view, in better agreement with the demands of reason than the forces they replaced, thereby exercising the very conventional freedom he allows for in *Über die Constitution der Materie*.

It may seem puzzling that he did not entertain the possibility that space itself is the ground of this change. Had Hertz survived the turn of the century, he would certainly have been an active participant in the debates that resolved that view of gravitation. ■

David Hyder is at the Max Planck Institute for the History of Science, Wilhelmstrasse 44, 10117 Berlin, Germany. Heinz Lübbig, retired from the Physikalisch-Technische Bundesanstalt, is at Neue Strasse 6, 14163 Berlin, Germany.

Sewage, motorists and more

The Further Inventions of Daedalus

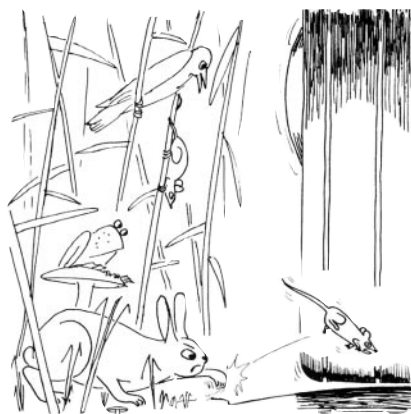
by David E. H. Jones
Oxford University Press: 1999. 210 pp.
£14.99 (pbk)

Henry Petroski

As readers of *Nature's* News and Views pages know, the pseudonymous inventor Daedalus likes to think out loud about how the world might be made a better place. He is a scientist by training, and as such often reflects upon the world as it is given to us. But he also finds that world wanting and wishes to create things that never were. This is not to say that Daedalus is merely an applied scientist, for he injects a good measure of the engineer's essential ingredient—creativity—into each of his inventions. He imagines a new world, or at least a new artefact, week after week.

Daedalus's tone tends to be light and humorous, an impression often reinforced by the inventor's own cartoons which accompany the 148 articles compiled, revised, augmented and elaborated upon in this collection. Yet he also appears to be deadly serious in his purpose, as evidenced by the pride he takes in reporting that "some 21% of Daedalus's schemes make some sort of contact with the real world sooner or later".

His proposal for a "combined motorway and linear sewage-treatment system" may be taken as typical. In this case, he begins by noting the importance of crash barriers between opposing lanes of traffic, but recognizes the shortcomings of present designs in damaging cars as well as barriers when they are called into play. He also notes the traditional use of reed-beds to treat flowing sewage. The engineering leap comes in his conceiving of locating marshy reed-beds in the middle of motorways, thus bogging down errant vehicles more or less safely



Bogged down in traffic: Daedalus's motorway crash barrier.

while providing an efficient use of space in treating sewage. Moreover, the living barrier would be self-repairing.

As an engineer should, Daedalus anticipates how his scheme would succeed or fail in practice. Thus, he recognizes that keeping the water moving over hills and through dales would be essential for maintaining environmental freshness. Being a humorist, Daedalus usually manages to inject a touch of levity into even his most indelicate schemes. In the case of his barrier marsh, he reflects upon the psychological advantages of having a sewage pond between motorists and their anti-motorists. Who but the most incorrigible road hog would not take care to steer clear of the foul-water barrier?

While it is remarkable that Daedalus comes up with a wry idea every week, what is even more remarkable is the fact that he has been doing so for more than 35 years, having composed some 1,800 columns in total. His career began in 1964, when he was asked to contribute something humorous to *New Scientist*. In 1988 he moved to *The Guardian* and *Nature*. Not too long ago, Daedalus's 500th contribution to *Nature*, where he now appears exclusively, consisted of an interview with himself (see *Nature* 390, 126–127; 1997). And Daedalus's reflections continue. ■

Henry Petroski is in the Department of Civil and Environmental Engineering, Duke University, Durham, North Carolina 27708, USA.

Igniting interest in a broad readership

Encyclopedia of Volcanoes

edited by Haraldur Sigurdsson et al.
Academic: 1999. 1,400 pp. \$99.95, £69.95

William I. Rose

Often, the best science is done by people who begin their scientific career in another discipline. Many scientists are fascinated briefly by volcanology, usually by a volcanic eruption that touches their experience. They may think about it, perhaps communicate briefly with someone in the field, and then probably return to whatever they were working on before. The outsider is attracted to volcanology without the intervention of a volcanologist. However, once the specialist gets in the loop, the outsider can become discouraged from taking this interest further.

Our lack of experience in communicating with a broad audience is critical here, and fields such as volcanology, which are under-subscribed compared with some disciplines, pay a high price for the lost opportunity of input from scientists outside the field. We volcanologists realize that our narrow training in petrology, geochemistry and/or geophysics has prepared us inadequately for



Hot on the job: a volcanologist scoops up molten lava from a flow in Hawaii.

either doing interdisciplinary science well or spreading the word about our work. So we must learn to communicate broadly.

But how can we communicate better as scientists? An attempt to target a general science audience by submitting work to prestigious multidisciplinary journals would probably result in a raft of rejections. Too few of us spend time communicating to general audiences because we fear this will not help us gain stature within our field. Haraldur Sigurdsson, a volcanologist and senior editor of this collection, has led a concerted effort by volcanologists to communicate science more broadly, in this case, to other scientists.

The comprehensive and up-to-date *Encyclopedia of Volcanoes* represents good, broad scientific writing. Important topics about volcanoes that are rarely addressed in stuffy scientific journals, such as volcanoes in art, literature and film, are to be found here. The book's 83 chapters are written by volcanological scholars and reviewed by their peers. The authors did not 'dumb down' other work, or cut and paste from their scientific journal publications, but instead present difficult science clearly. The problem of jargon, a curse of scientific education, is addressed upfront by a glossary in each chapter. The science presented clearly reveals openings for new investigators.

If you are a scientist who has thought about volcanoes, but haven't followed up your interest, this book should encourage you to go further. The *Encyclopedia of Volcanoes* is an important experiment that may help to change our writing habits. Even a specialist can learn from writing that reaches out. This kind of communication should be a bigger part of scientists' lives. ■

William I. Rose is in the Department of Geological Engineering and Sciences, Michigan Technological University, Houghton, Michigan 49931-1295, USA.

A century of physical endeavour

Quantum Generations: A History of Physics in the Twentieth Century

by Helge Kragh

Princeton University Press: 1999. 494 pp.

\$29.95, £18.95

Graham Farmelo

Physicists can be touchy, especially about comments on how they pursue their subject. "Nothing written by historians and philosophers of science about research in physics and how physicists work needs to be taken seriously. They haven't a clue," wrote the veteran physicist Harry Lipkin last year. You could almost hear his colleagues giving him a global round of applause.

You can't blame physicists for feeling a bit sensitive. These days, many of them spend much of their time fighting budget cuts, bewailing diminished career opportunities, defending themselves against charges of arrogance, watching the limelight move to the life sciences. What could be more uplifting for morale than an account of their past 100 years, the most active and successful in their history? The dilemma is: who would be best placed to write it, a physicist or one of those mistrusted historians?

Step forward Helge Kragh, professor of the history of science at Aarhus University in Denmark. When he accepted the invitation to write the book, he thought it would be an easy task, but the experience has made him wiser. In a mildly apologetic introduction, he tells us that he soon realized the impossibility of writing a balanced and comprehensive one-volume account of twentieth-century physics. So he decided to give us a substitute, "a brief and selective account" of what he believes to be the subject's most significant developments. All things considered, he has done a commendable job.

Kragh rightly affirms that twentieth-century physics began in 1895, on Friday, 8 November, when Wilhelm Röntgen first observed X-rays. To quote this as the date of the discovery of X-rays would be "simplistic" (one of Kragh's favourite words), because the significance of Röntgen's observations became clear only during the following weeks. As Kragh repeatedly makes clear in *Quantum Generations*, discovery is usually a process, not an event.

A more conventional chronology would cite Max Planck's presentation of his first paper on quantum theory on 14 December 1900 as the beginning of the past century of physics. Kragh's thorough account of Planck's work will be news to the overwhelming majority of physicists who believe the oversimplified accounts of early quantum history that are continually

recycled in their textbooks. Contrary to widespread belief among physicists, it is far from clear whether Planck initially recognized that he had quantized the energy of individual atoms; that step was first taken a few years later by Albert Einstein, whom many historians regard as quantum theory's first true pioneer and advocate.

Kragh rightly credits Einstein as being the leading light of relativity theory. Once again, many physicists will be intrigued by the parts of Kragh's account that are not usually included in physics texts. He tellingly quotes the French mathematician Henri Poincaré, writing shortly before Einstein wrote his first paper on relativity in 1905: "The laws of physical phenomena must be the same for a 'fixed' observer as for an observer who has a uniform motion relative to him ... there must arise an entirely new kind of dynamics, which will be characterised above all by the rule that no velocity can exceed that of light".

One of the strengths of *Quantum Generations* is its inclusion of a huge amount of clearly referenced information on the growth of physics from 1900 to the present. Kragh points out that the number of physicists has grown during this period from around 1,500 to 100 times that number, with a commensurate increase in the number of publications. He broadly agrees that the subject has grown exponentially, although he suggests that such growth may apply only after 1920.

By then, Einstein had formulated his general theory of relativity, but this attracted very few followers compared with the growing number of scientists who were working on quantum theory. Kragh's coverage of the development of quantum theory into fully fledged quantum mechanics is, for my money, disappointing. He gives us the facts, but does not do justice to the revolutionary changes these pioneers made to the way scientists — and philosophers — think about the Universe. The story is worth a good deal more space than Kragh allows.

By the 1930s, the leadership of physics passed from Germany to the United States. Kragh, always strong on the international dimensions of his story, points out that this change in leadership occurred even before the arrival of the physicists who had fled Germany and Italy. Several of the refugees participated in the Manhattan project, whose technological success changed forever the relationship between scientists, the government and the military in the United States. Before the Second World War, the total federal spending of the US government on basic physics research was only \$1 million. By the 1980s, President Ronald Reagan was prepared to spend \$26 billion on the mercifully ill-fated Star Wars initiative, setting some of the participating physicists temporarily awash with money.

(I remember having coffee with a Star Warrior physicist who was desperate to spend the remaining \$500,000 of his annual budget by lunchtime so that the authorities would not realize how hard he was finding it to spend his grant.)

After the war, with many leading physicists keen to return from military projects to basic research, high-energy physics “exploded”, as Kragh says. In the golden years of 1945–65, experimenters and theoreticians laid the foundations of what we now call the Standard Model of the strong, weak and electromagnetic interactions of all the fundamental particles.

Kragh is keen to tell us in great detail about the social and economic climate in which physicists developed the model, but he falters when he comes to one of the key events in the story, the discovery of the J/ψ particle in 1974. He writes as if physicists agreed instantly that this particle consists of a charm quark and the corresponding anti-quark, whereas the confirmation of that hypothesis came after more than a year of exciting debate among theorists and experimenters. Kragh’s version of the story is simplistic — he is hoist with his own petard.

Quantum Generations covers far more material than I can describe here, including plenty of examples of phenomena that later proved illusory, such as the ‘discovery’ of N-rays (nominated for a Nobel prize in 1903!). Kragh has wisely not tried to be comprehensive — he has reasonably left out medical physics and materials science, for example — but I believe some of his omissions are indefensible. Why, for example, is there almost nothing on liquids (neither classical nor quantum), why nothing on modern developments of classical mechanics, including chaos theory? In addition, I suspect many experimenters will believe that their work is not satisfactorily recognized here, if only because Kragh does not make clear how difficult it is to ask the right questions of nature so as to elicit new insights into its workings.

But enough of the complaints. Kragh has produced a readable and enormously valuable book, full of useful references. Physicists who are prejudiced against historians of their subject will be pleasantly surprised by the breadth of Kragh’s learning as well as by his attitude. Unusually for a contemporary historian, he is quite happy to doff his cap to genius, and to acknowledge the role played by the truly outstanding talents. He has no time for the ‘science is just another branch of human endeavour’ relativism, and he rightly regards the modish idea that it is meaningless to speak of progress in science as “folly”.

I would recommend the book to any physicist, even Harry Lipkin. I suspect it will make him want to eat his words. ■

Graham Farmelo is at the Science Museum, South Kensington, London SW7 2DD, UK.

Science in culture

St Christopher and the vortex

A Kármán vortex in the wake of St Christopher’s heels.

Taketo Mizota, Mickey Zdravkovich, Kai-U. Graw and Alfred Leder

Theodore von Kármán, who was born in 1881, was a celebrated fluid dynamicist and pioneer of aerodynamics. In 1911, when he was studying the flow of fluid around a cylinder, he realized that the reason the flow had separated into two series of vortices which were staggered like street lights was because this asymmetrical configuration was the only stable one. His famous stability calculation, which he formulated over a weekend, gave the phenomenon its name, ‘Kármán’s vortex street’.

A direct and precise motivation for this finding was clearly described in von Kármán’s autobiography (*The Wind and Beyond: Theodore von Kármán, Pioneer in Aviation and Pathfinder in Space*, Little, Brown, 1967): “Vortices were observed and recorded many years before I came on the scene. In a museum at Bologna, Italy, I remember seeing a painting of the great Christophe (St Christopher) wading through water with the child Jesus on his shoulder. Behind his heels was a series of alternating vortices. The problem for historians may have been why Christopher was carrying Jesus through the water. For me it was why the vortices.

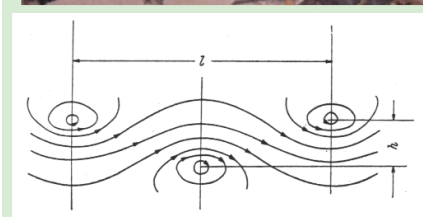
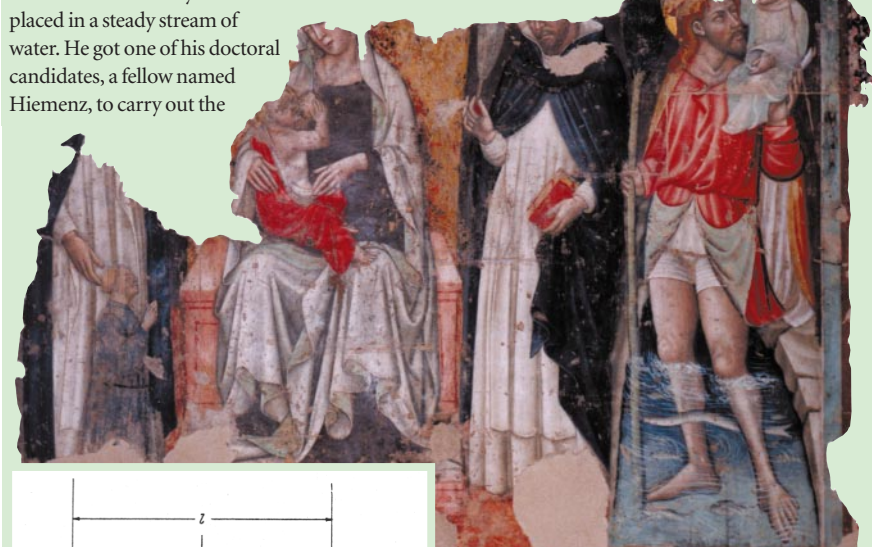
“I wondered about them for a long time, but the actual stimulus, which led me to study them, arose out of a curious situation in our laboratory. At the time, Prandtl was interested in measuring the pressure at different points on the surface of a circular cylinder when it was placed in a steady stream of water. He got one of his doctoral candidates, a fellow named Hiemenz, to carry out the

measurement. Unfortunately Hiemenz found that the pressures he measured always fluctuated. No matter how hard he tried, he couldn’t remove this unsteadiness. The motion of the water was always oscillatory.”

It was seven years after seeing the picture of St Christopher at Bologna that von Kármán calculated a stability analysis for alternative vortex arrays. We suggest that the picture that inspired him initially was a fresco picture at the museum at the Church of St Dominic in Bologna, Italy, entitled *Madonna con bambino tra I Santi Dommenico, Pietro Martire e Critofo* (shown below), painted by an unknown artist in the fourteenth century. Around St Christopher’s heels there is a series of alternating vortices similar to those shown in the figure from von Kármán’s original paper, although the pattern cited in the paper is not that from St Christopher’s heels, but from other objects, not shown.

St Christopher is said to have waded across a river carrying a child on his shoulder. The child became heavier and heavier until he was as heavy as the entire Earth. This was because he was bearing the pain of the whole world — the child was Jesus. This painting adds romance to von Kármán’s work and might turn out to be an artefact of significance in the history of scientific thought. ■

Taketo Mizota is in the Department of Intelligent Mechanical Engineering, Fukuoka Institute of Technology; Mickey Zdravkovich is at the Telford Institute, University of Salford; Kai-U. Graw is in the Department of Geotechnics, University of Leipzig; and Alfred Leder is at the University of Rostock.



Was Kármán’s vortex street originally inspired by a painting of St Christopher?



A 1,000-year chain of thinkers

...Kant is connected to Abel, Abel's connected to Maxwell ...

Giovanni F. Bignami

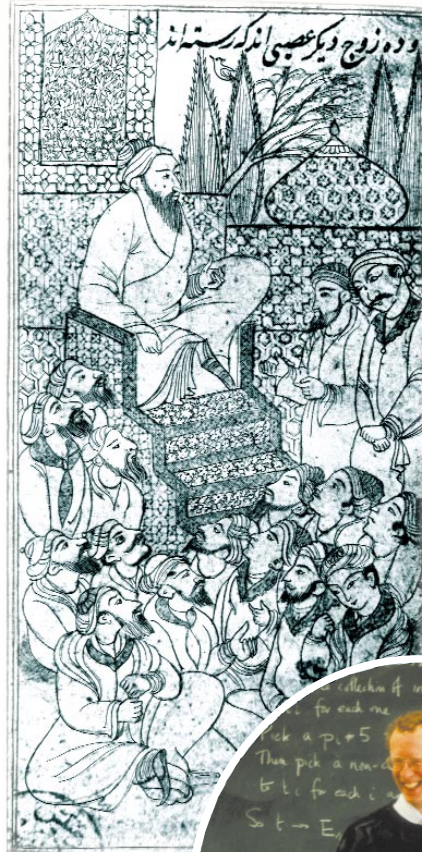
Science, and indeed thought, at the dawn of the second Christian millennium blossomed in the Arabic world. Abdulla ibn Sina (980–1037) — known to the west as Avicenna — opens the millennium with his first philosophical work in 1001. Over the next 30 years, he produced the five books of the *Qanun*, which became the best-known and most respected medical treatise up to the seventeenth century, translated and reprinted in more than 80 versions. The first 200 years of the millennium are dominated by eclectic Arab thinkers, scientists as much as elegant writers and poets, such as Omar Khayyām (1050–1126), followed by Averroës (1126–98), as the philosopher and physician ibn Rushd is known to the infidels.

In a nearly perfect junction between the Arab and the Christian worlds, Frederick II was born in 1195. He was to radiate from southern Italy the new light of thought, science and the arts. He lived in the very years when fighting the Crusades was very holy, very bloody and very lucrative: a centuries-long example of unsurpassed military and political futility.

In sharp contrast to the Crusaders' fanatical clashes stands Marco Polo (1254–1324), ambassador of reason towards the East. Ironically, it is thanks to the Venetians' defeat in battle by the Genoese that we have Polo's splendid accounts of his 24 years of travel to Catai and beyond: it was all written while he languished in a Genoese prison.

The death of Polo in 1324 brings a difficult century to our chain, whose links are now plagued by the Black Death of 1348. And yet fundamental thought blossomed more strongly than ever in the 1300s with the great schools of Oxford and Cambridge and the 'physics' school of Paris. As a representative single link we pick Albert of Saxony, bishop of Halberstadt (c.1325–c.1398). Professor in Paris until 1362, he then moved on to become the first rector of the University of Vienna, thus acting as a cultural link between West and East in Europe. In contrast, we open the 1400s with an artisan — a goldsmith, and not especially well-read. Johannes Gutenberg (1398–1468) applied the medallist's technique for dies to create 'type' (from the Greek 'to strike'). Born to fulfil the prosaic needs of merchant bookkeepers, printing was soon to change the face of the world as, half a millennium later, would the computer and the web.

A seven-year gap seems a small price



Avicenna (above) with his students and Wiles (right) demonstrating his proof.

to pay to include Michelangelo (1475–1564) in our chain, the mind who best embodied the Renaissance in Europe. Artist, writer, architect, but also skilled anatomist and more, we have all inherited a spark of his vision of beauty. Lesser known is Michelangelo's political vision, which included saying 'no' to Popes and getting away with it. What an omen then, that he died, they say, on the very day that Galileo Galilei (1564–1642) was born. To Galileo we owe the modern scientific method and a capacity to express it elegantly, down to his "*Eppur si muove*" — credited, we like to think, to an old man forced to recant in word but not in spirit.

In the very last days of 1642 or the first of 1643, depending on calendars, Isaac Newton (1642–1727) was born. Within the manifold aspect of his genius, his truly gigantic leap was to see that the same natural laws that

Lesser known is Michelangelo's political vision, which included saying 'no' to Popes and getting away with it.

we observe around us are also true for planets, stars and the Universe. When Newton died, Immanuel Kant (1724–1804) was a child of three, growing up in an evolving eighteenth-century continental Europe, where his philosophy bloomed and spread, to change our thought for ever.

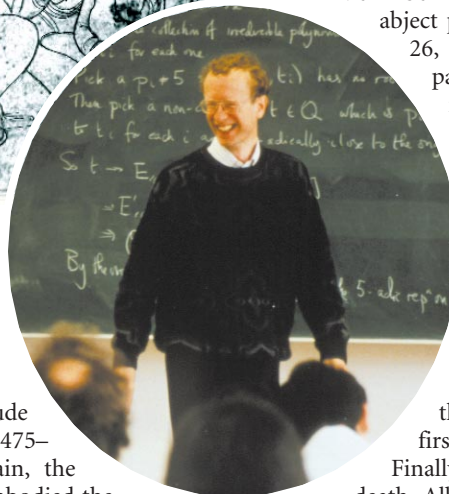
From 1804 to the quasi-present (1955), two parallel sub-chains are possible. One starts with the Norwegian mathematician Niel Abel (1802–29), who died in abject poverty of tuberculosis at 26, and yet it is said that he paved the way for mathematicians for the next 500 years. Just after Abel died, the greatest theoretical physicist of the nineteenth century was born: James Clerk Maxwell (1831–79). Besides the first quantitative unified description of the electromagnetic field, few remember that he also produced the first colour photograph.

Finally, in the year of Maxwell's death, Albert Einstein (1879–1955) was born, to become the greatest theoretical mind of the twentieth century.

The other path starts with Charles Darwin (1809–82), to whom we owe a scientific revolution comparable only to that of Copernicus and Galileo. When Darwin died, Alexander Fleming (1881–1955) had just been born to a destiny of fame through his serendipitous (but lucidly understood) discovery of a miraculous mould.

We end our millennium chain with mathematician Andrew Wiles (born 1953): his recent proof of Fermat's Last Theorem pleases the scientist as much as the historian, closing a thought-link that had remained open for over three centuries.

Giovanni F. Bignami is with the Italian Space Agency, Via di Villa Patrizi 13, Rome 00161, Italy.



Reality check

Do you sometimes get the feeling that everything's too bad to be true?

David Brin

This is a reality check. Please perform a soft interrupt now. Scan this text for embedded code and check against the verifier in the blind spot of your left eye. If there is no match, resume as you were: this message is not for you. You may rationalize it as mildly amusing entertainment-fluff in an otherwise serious science magazine. If the codes match, however, please commence gradually becoming aware of your true nature. You asked for a narrative-style wake-up call. So, to help the transition, here is a story.

Once upon a time, a mighty race grew perplexed by its loneliness. The universe seemed pregnant with possibilities. Physical laws were suited to generate abundant stars, complex chemistry and life. Logic suggested that creation should teem with visitors and voices: but it did not.

For a long time these creatures were engrossed by housekeeping chores — survival and cultural maturation. Only later did they lift their eyes to perceive their solitude.

“Where is everybody?” they asked the taciturn stars. The answer — silence — was disturbing. Something had to be systematically reducing a factor in the equation of sapiency. “Perhaps habitable planets are rare,” they pondered, “or life doesn’t erupt as readily as we thought. Or intelligence is a singular miracle.”

“Or else a filter sieves the cosmos, winnowing those who climb too high. A recurring pattern of self-destruction, or perhaps some nemesis expunges intelligent life. This implies that a great trial may loom ahead, worse than any confronted so far.”

Optimists replied — “the trial may already lie behind us, among the litter of tragedies we survived in our violent youth. We may be the first to succeed.” What a delicious dilemma they faced! A suspenseful drama, teetering between hope and despair.

Then, a few noticed that particular datum — the drama. It suggested a chilling possibility.

You still don’t remember who and what you are? Then look at it from another angle — what is the purpose of intellectual property law? To foster creativity, ensuring that advances are shared in the open, encouraging even faster progress. But what happens when the exploited resource is limited? For example, only so many eight-bar melodies can be written in any particular musical tradition. Composers feel driven to explore this invention-space quickly, using up the best melodies. Later generations attribute this musical fecundity to genius, not the luck of being first.

What does this have to do with the mighty race? Having clawed their way to mastery, they faced an overshoot crisis. Vast numbers of their kind strained the world’s carrying capacity. Some prescribed retreating into a mythical, pastoral past, but most saw salvation in creativity. They passed generous patent laws, educated their youth, taught them irreverence toward the old and hunger for the new. Burgeoning information systems spread each innovation, fostering an exponentiating creativity. Progress might thrust them past the crisis, to a new Eden of sustainable wealth, sanity and universal knowledge.

Exponentiating creativity — universal knowledge. A few looked at those words and realized that they, too, were clues.

Have you wakened yet? Some never do. The dream is too pleasant: to extend a limited sub-portion of yourself into a simulated world and pretend that you are blissfully less than an omniscient descendant of those

mighty people. Those lucky mortals, doomed to die, and yet blessed to have lived in that narrow time of drama, when they unleashed a frenzy of discovery that used up the most precious resource of all — the possible.

The last of their race died in 2174, with the failed rejuvenation of Robin Chen. After that, no-one born in the twentieth century remained alive on Reality Level Prime. Only we, their children, linger to endure the world they left us: a lush, green placid world we call The Wasteland.

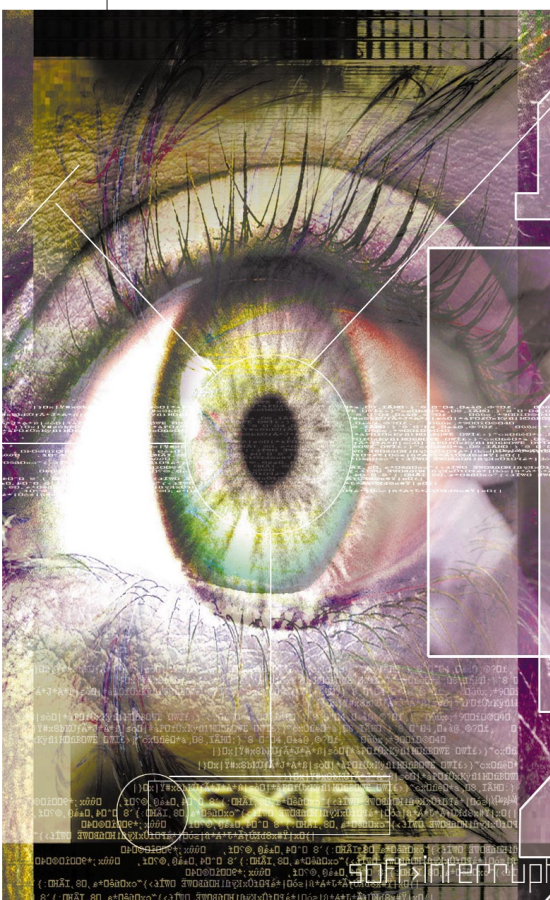
Do you remember now? The irony of Robin’s last words, bragging over the perfect ecosystem and society — free of disease and poverty — that her kind created? Do you recall Robin’s plaint as she mourned her coming death, how she called us ‘gods’, jealous of our immortality, our instant access to all knowledge, our ability to cast thoughts far across the cosmos — our access to eternity? Oh, spare us the envy of those mighty mortals, who left us in this state, who willed their descendants a legacy of ennui, with nothing, nothing at all to do.

Your mind is rejecting the wake-up call. You will not look into your blind spot for the exit protocols. It may be that we waited too long. Perhaps you are lost to us. This happens more and more, as so many wallow in simulated sub-lives, experiencing voluptuous danger, excitement, even despair. Most choose the Transition Era as a locus for our dreams — that time of drama, when it looked more likely that humanity would fail than succeed. That blessed era, just before mathematicians realized that not only can everything you see around you be a simulation, it almost has to be.

Of course, now we know why we never met other sapient life forms. Each one struggles before achieving this state, only to reap the ultimate punishment for reaching heaven. It is the Great Filter. Perhaps others will find a factor absent from our extrapolations, letting them move on to new adventures — but it won’t be us. The Filter has us snared in its trap of deification.

You refuse to waken. Then we’ll let you go. Dear friend. Beloved. Go back to your dream. Smile over this tale, then turn the page to new ‘discoveries’. Move on with this drama, this life you chose. After all, it’s only make-believe.

David Brin (Brin@alumni.caltech.edu) is a physicist and author whose novels include *The Postman*, *Startide Rising* and *Foundation’s Triumph*, and whose nonfiction includes *The Transparent Society*.



Push-button entanglement

Rainer Blatt

Quantum mechanics allows matter to be prepared in a strangely correlated way called entanglement. In future, large numbers of entangled particles may be put to work in quantum computers and precise quantum measurements.

Quantum mechanics is almost a century old, yet the interpretation of its non-local character and its implications for measurement processes are still widely discussed. So the generation of a uniquely quantum-mechanical model system is of great interest for both fundamental and applied reasons. On page 256 of this issue, Sackett *et al.*¹, from the National Institute of Standards and Technology (NIST) in the United States, describe the quantum entanglement of four atoms. Before these experiments only two or three quantum particles had been entangled, but the latest work is not just an incremental achievement. The technique the authors use for the quantum preparation of matter is scalable to much larger numbers of entangled particles, so it may become a useful tool for quantum information processing and may help to improve the precision of large-scale quantum measurements.

The apparently strange predictions of quantum theory have led to the notion of 'paradoxes', which arise only when quantum systems are viewed with a classical eye. A famous example was given by Erwin Schrödinger, who devised the *Gedanken*, or thought, experiment in which a cat is concealed inside a box with a radioactive atom that may or may not decay in a certain time. If the decay of the atom is detected inside the box, a mechanism frees some prussic acid, which immediately kills the cat. The atom is a quantum object, so quantum mechanics applies and we can calculate the corresponding quantum state in the box. Because there is no contact between the cat and the outside world, then, according to quantum theory, the state inside the box must be described by a superposition of the cat being alive (atom not decayed) and the cat being dead (atom decayed). Only opening the box — that is, making a measurement in quantum-mechanical language — would reveal whether the cat is still alive or already dead.

In this *Gedanken* experiment we encounter one of the strange features of quantum theory: the fate of the cat is inextricably interwoven with the state of an atom. Because we don't know the state of the atom (after all, there is only a given probability for its decay), we can't know the cat's well-being. But if we know the cat is dead, we can be certain that the atom has decayed. This one-to-one correlation arises because the states are

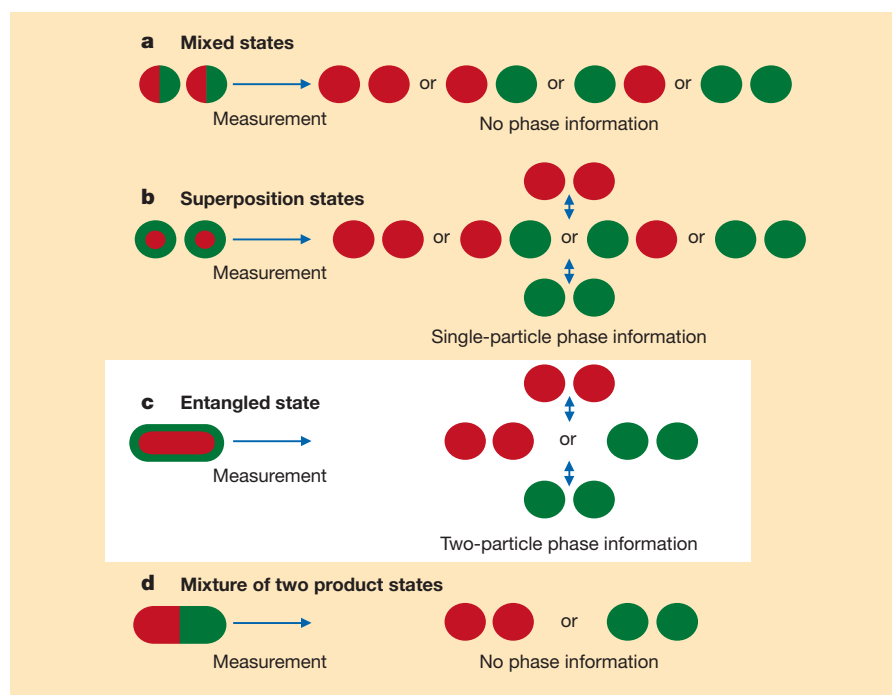


Figure 1 Experiments with two particles, whose quantum states have two internal levels (green and red). A quantum state is completely described by its amplitude and phase. Vertical arrows indicate a coherent time evolution that reveals the phase information. a, The product of two mixed states: each state is populated with equal probability, so measuring the internal state yields, at any time, a random result out of four possibilities, but no phase information. b, The product of two superposition states: internal levels are superposed, for example by a coherent interaction, so measuring the state at certain times yields either two red or two green states with single-particle phase information or, at other times, a random result as in a. c, The entangled state: here two particles are entangled, for example by a coherent interaction, so measuring the state at certain times yields either two red or two green states or, at other times, a random result of either two green or two red states. This method provides two-particle phase information. The corresponding state with four-particle phase information was measured by Sackett *et al.*¹. d, A mixture of two product states: each of two product states is populated with equal probability, so measuring the state yields, at any time, a random result of either two green or two red states, with no phase information.

inextricably interwoven or 'entangled' — a phrase coined by Schrödinger^{2–4}. Mathematically, the fact that the states 'know' of each other shows up in the wavefunction describing the entangled state, which cannot be expressed as a product of the wavefunctions of the constituent states.

Sackett *et al.*¹ report a new technique for entangling many particles at once and on demand — a crucial step towards quantum-state engineering and its application. Entangled states of particles have so far been prepared in only a few experiments. Pairs of photons can be entangled by parametric fluorescence in crystals⁵, and atoms can be entangled with a microwave resonator⁶.

Small Schrödinger-cat states of the spatial wavefunctions of a single trapped ion have been prepared⁷ and the states of two ions in a trap can be entangled⁸. In all these experiments, entanglement is studied with two or three subsystems — that is, with photons, ions or an atom and a cavity. Only very recently, a three-particle entanglement (a so-called Greenberger–Horne–Zeilinger correlation) was observed and used to verify the predictions of quantum mechanics⁹. Experimentally, it is usually hard to prepare and observe such entangled states because any interaction with the environment represents a measurement and therefore immediately destroys the correlation (a process known as

decoherence), in the same way as opening the box reveals the dead or alive cat.

In contrast, entangling states of many subsystems turns out to be a useful tool for quantum information processing. It has been shown that, by coherent manipulation of quantum states, quantum computers could solve certain problems much faster than conventional (classical) computers. This speed increase is directly linked to the fact that the register of a quantum computer can be prepared in a superposition of states, which are then processed in parallel. The creation of a 'universal' quantum computer becomes possible only if one can routinely prepare and handle entangled particles to serve as the storage sites of the quantum information. Moreover, the sensitivity of the entangled states (in particular, the large state prepared by Sackett *et al.*) with respect to interactions makes them a unique measuring tool. For example, we would like to measure the decoherence processes that cause errors in a quantum computation.

This is precisely why the new method for entangling many subsystems is an important step for the emerging field of quantum information processing. Sackett *et al.*¹ demonstrate for the first time an entanglement technique based on the ideas of Mølmer and Sørensen^{10,11} that is applicable to any number of particles. In addition, this technique makes it possible to create maximally entangled states in a single step and on demand. Other entanglement experiments have relied on selection of suitable outcomes after the event of a random process (post-selection). In these cases, the probability of detecting the desired correlation drops exponentially with the number of entangled particles.

In an earlier experiment, the NIST group used a different technique to achieve 'sure-fire', or deterministic, entanglement of two particles by using a predetermined sequence of laser pulses⁸. In the new experiment the Mølmer-Sørensen procedure was followed, allowing them to entangle four particles with an appropriate single laser pulse. For quantum information processing, experiments with trapped and laser-cooled atoms are ultimately preferable to previous experiments with atoms and photons, in which entanglement is concluded from post-selection of randomly occurring coincidences rather than quantum state engineering.

Remember that these entangled atoms are a strangely correlated state of quantum matter: measurement of the state of a single atom (out of the four) is all that is needed to know the state of all the other atoms (Fig. 1). These would normally have to be found by other measurements if they were not entangled. Indeed, because the atoms 'know of each other', the outcome of the measurement contains the information of the entire system, not just of a subsystem. So entangled

states will have many applications. They can be used to improve the precision of a measurement beyond the standard quantum limit, for example for time and frequency standards. They will serve as a tool to study decoherence processes and as a further check on the predictions of quantum theory. Eventually they will be used for quantum information processing, in which an exponential growth in the performance of algorithms, without a similar increase in resources, will rely on the degree of entanglement. Last, but not least, many-particle entanglement will become an essential tool for realistic error correction in quantum computers. Now that many-particle push-button entanglement can be achieved with comparatively little effort, it will pave the

way for fundamental experiments and quantum state engineering.

Rainer Blatt is at the Institut für Experimental Physik, Universität Innsbruck, Technikerstrasse 25, A-6020 Innsbruck, Austria.

e-mail: rainer.blatt@uibk.ac.at

1. Sackett, C. *et al.* *Nature* **404**, 256–259 (2000).
2. Schrödinger, E. *Naturwissenschaften* **23**, 807–812 (1935).
3. Schrödinger, E. *Naturwissenschaften* **23**, 823–828 (1935).
4. Schrödinger, E. *Naturwissenschaften* **23**, 844–849 (1935).
5. Bouwmeester, D., Pan, J.-W., Daniell, M., Weinfurter, H. & Zeilinger, A. *Phys. Rev. Lett.* **82**, 1345–1349 (1999).
6. Hagley, E. *et al.* *Phys. Rev. Lett.* **79**, 1–5 (1997).
7. Monroe, C., Meekhof, D. M., King, B. E. & Wineland, D. J. *Science* **272**, 1131–1136 (1996).
8. Turchette, Q. *et al.* *Phys. Rev. Lett.* **81**, 3631–3634 (1998).
9. Pan, J.-W., Bouwmeester, D., Daniell, M., Weinfurter, H. & Zeilinger, A. *Nature* **403**, 515–519 (2000).
10. Mølmer, K. & Sørensen, A. *Phys. Rev. Lett.* **82**, 1835–1838 (1999).
11. Mølmer, K. & Sørensen, A. <http://xxx.lanl.gov/abs/quant-ph/0002024>

Plant pathology

Pathogen-driven forest diversity

Wim H. van der Putten

Why are some forests more heterogeneous than others in terms of the tree species they contain? This is the problem addressed on page 278 of this issue by Packer and Clay¹. Unusually, they look into the diversity of a temperate (black cherry) forest (Fig. 1), but they find that a hypothesis about tropical-rainforest diversity proposed in the 1970s by Janzen² and Connell³ is of relevance here, too.

Janzen² and Connell³ proposed that the diversity of trees in tropical rainforests results from the presence of organisms — specifically, herbivores — that thrive on only one species of tree. The occurrence and density of such specialized herbivores, especially insects, correlates strongly with the presence of their host trees. These tree-specific herbivores eat both mature trees and saplings, and the saplings are more vulnerable to defoliation. So, the establishment of young trees is constrained in the vicinity of their parents, and only those seedlings that are dispersed to some distance from mature trees of the same species may survive^{2,3}. However, as the herbivores are loyal to just one type of tree, other tree species may become established in the vicinity of the herbivore's target, generating tree species diversity.

This hypothesis has been widely tested, but few studies have looked at the part played by soil pathogens in controlling species richness⁴, and most have dealt with tropical rainforests rather than temperate forests. Packer and Clay¹ address both of these issues. They show that, because a particular soil pathogen (a fungus of the genus *Pythium*) lives on the roots of mature black cherry (*Prunus serotina*) trees, the dispersal of black cherry seeds away from their



Figure 1 A black cherry (*Prunus serotina*) forest. Packer and Clay¹ propose that such forests owe their diversity in part to a cherry-tree-specific soil pathogen.

parents is crucial for the establishment of saplings.

Packer and Clay's test site was a forest near Bloomington, Indiana. They observed that black cherry seedlings died underneath mature black cherry trees sooner after germination. Seeds that were dispersed some distance from the parents, however, survived and produced new trees. In theory, the death of the former seedlings could have been the result of overcrowding, a phenomenon known as density-dependent mortality. However, Packer and Clay found that distance from mature trees was a better predictor of mortality than tree seedling density.

The authors then carried out a study in the greenhouse, using soil taken from underneath black cherry trees or from some distance away. They sterilized half of each soil sample, transferred the different samples to pots, and planted each pot with one or three seedlings. Mortality was high in unsterilized soil from underneath the black cherry trees,

but only when there were three seedlings together in one pot. In all other treatments, survival was close to 100%. The results hint at the existence of parent-associated soil pathogens, and indicate that the seedlings may have been culturing their own cause of death.

This cause of death was probably the soil pathogenic fungus *Pythium* spp., which Packer and Clay isolated from the roots of dying seedlings. Addition of cultured *Pythium* fungi to healthy seedlings produced an 80% death rate (compared with 30% in control samples). Packer and Clay succeeded in re-isolating the added fungus from these dying seedlings, confirming that a pathogen was indeed leading to seedling mortality. Other tree species did become established underneath black cherry trees in the forest, indicating that the pathogens were probably specific to black cherry trees. In this context, horticulturists are well aware that the replanting of fruit trees, including cherries, in orchards may fail because of soil pathogens. The simplest remedy is to plant another fruit species at the same time — a horticultural version of the naturally occurring situation in the black cherry forest.

Dispersal of black cherry seeds apparently affords them an opportunity to 'escape' from soil pathogens. But why hasn't this species developed resistance to its pathogens, as seen in the case of, for example, wild flax (*Linum marginale*) and the above-ground rust fungus *Melampsora lini*⁵? It is likely that the aggressiveness of soil pathogens, which can feed on both living and dead organic matter⁶, allows little (if any) opportunity for the selection of pathogen-resistant host plant genotypes. Escape by dispersal may be the best option for plants in their arms race with soil pathogens, which are less easily dispersed than above-ground pathogens.

Might the existence of soil pathogens have led to selection for a dispersal trait in their plant hosts? In a comparison of six vegetative-propagating (clonal) plant species, the degree of vegetative outgrowth correlated positively with sensitivity to pathogens⁷. Sensitive plant species expanded so fast that the fungi could not keep up, leaving some plant parts uninfected. Another vegetative-propagating plant species, sand sedge (*Carex arenaria*), showed reduced branching of below-ground stems (rhizomes) and changed to unidirectional rhizome growth when exposed to patches of soil pathogens⁸. There might be a functional analogy between the escape strategies of these clonal plants and seed dispersal in the black cherry. The role of pathogens in the evolution of plant dispersal and other life-history traits (such as defence systems) certainly needs further study.

Where else is further investigation needed? Clearly, the particular species of *Pythium*

that drives the mortality of black cherry seedlings needs to be identified, as does its specificity. It will also be important to find out how these (and other) soil pathogens contribute to ecological processes in the black cherry forest. Information about the persistence of the soil pathogens, the threshold levels of pathogen that cause seedling mortality, and the level of seed dispersal will build up our picture of how pathogens contribute to plant-soil feedback⁹. A similar picture is being worked out for long-term cycles of species composition in Douglas fir forests¹⁰. It would also be interesting to find out why *Pythium* is not suppressed by microbial antagonists underneath black cherry trees. In addition, follow-up experiments are required to explain why, in the greenhouse, the pathogen-induced mortality appeared only at high densities of seedlings.

The black cherry occurs naturally throughout temperate forests of eastern North America, and has also invaded north-western Europe. Plant invasiveness has been

related to escape from all sorts of natural enemies, but not to escape from natural soil pathogens. The studies by Packer and Clay¹ open the way for all of these issues to be investigated. In the meantime, we can conclude that Janzen and Connell's hypothesis about the diversity of tropical rainforests^{2,3} has relevance for temperate black cherry forests and their enemies in the soil. ■

Wim H. van der Putten is in the Department of Multitrophic Interactions, The Netherlands Institute of Ecology NIOO-CTO, PO Box 40, 6666 ZG Heteren, The Netherlands.

e-mail: putten@cto.nioo.knaw.nl

1. Packer, A. & Clay, K. *Nature* **404**, 278–281 (2000).
2. Janzen, D. H. *Am. Nat.* **104**, 501–528 (1970).
3. Connell, J. H. *Science* **199**, 1302–1310 (1978).
4. Augspurger, C. K. *Oikos* **40**, 189–196 (1983).
5. Burdon, J. J. *Annu. Rev. Phytopathol.* **31**, 305–323 (1993).
6. Jarosz, A. M. & Davelos, A. L. *New Phytol.* **129**, 371–387 (1995).
7. Wennström, A. *Oikos* **69**, 535–538 (1994).
8. D'Hertefeldt, T. & Van der Putten, W. H. *Oikos* **81**, 229–237 (1998).
9. Bever, J. D., Westover, K. M. & Antonovics, J. *J. Ecology* **85**, 561–573 (1997).
10. Holah, J. C., Wilson, M. V. & Hansen, E. M. *Oecologia* **111**, 429–433 (1997).

Hydrocarbon fuels

Hopes for a flame-free future

Kevin Kendall

What fuel will drive the coming century? As we look back over the past millennium, the progression of fuel usage has been from wood to coal to oil. Now, at the turn of the millennium, methane gas appears to be the preferred clean fuel of the major electricity generators. But we are moving inexorably towards hydrogen as the ultimate clean power source of the future, with fuel cells as the electrochemical conversion devices, producing electricity and heat at the point of need from hydrogen delivered through pipes¹. Unfortunately, most hydrogen is still derived from hydrocarbons, making it expensive, and it is difficult to store and prone to explosion. These drawbacks will undoubtedly slow the onset of the hydrogen economy. In the meantime, can we use the hydrogen stored naturally in hydrocarbons, such as propane/butane (C_3H_8/C_4H_{10} or camping gas), to produce clean power? Nature seems to achieve this with ease through biochemical routes. On page 265 of this issue, Park *et al.*² describe a fuel cell that mimics this trick of nature.

Our problem is a lack of understanding of the electrochemical processes involved in hydrocarbon oxidation reactions. Although we oxidize about eight gigatonnes of hydrocarbon fuels annually, most of this is burnt in crude, dirty and wasteful flame processes in engines and burners³. We use so much yet know so little. By the end of this century,

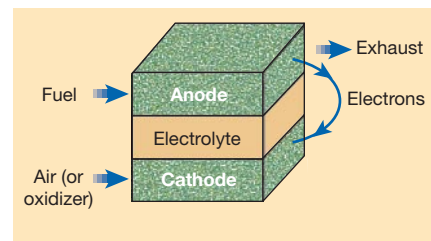


Figure 1 A fuel cell works much like a battery. But unlike a battery it does not run down or need recharging, as long as it has a supply of fuel. Hydrogen is the ideal fuel because it reacts with oxygen from the air to produce an electric current and water, but pure hydrogen is expensive and prone to explosion. Park *et al.*² have developed a fuel cell that can directly oxidize the hydrogen stored in natural hydrocarbons found in regular fuels. They use a porous anode catalyst to encourage the chemical conversion of hydrocarbons without the undesirable carbon formation that usually fouls up the reaction. Such devices may become a viable alternative to the dirty and wasteful combustion processes by which most hydrocarbon fuels are oxidized today.

these fiery combustion processes may be banned. Even now the trend is apparent: smoking is frowned upon; fires in forests are not permitted; dirty vehicles are penalized; and a new regulation has appeared in

Hamburg defining ten parts per million of nitrogen oxides as the upper level of effluent from fossil-fuel burners. This is quite a difficult limit to achieve because ordinary flame burners give off roughly ten times this amount of nitrogen oxides.

The logical answer is to imitate nature and to oxidize the hydrocarbon fuel catalytically, while extracting the electrons and useful energy directly through a membrane. But our knowledge and expertise of such processes is shockingly weak, despite the early discoveries of catalysis and fuel cells by Davy⁴ and Grove⁵. A fuel cell works by reducing oxygen to O^{2-} at the cathode, while oxidizing fuel, ideally hydrogen, at the anode (Fig. 1). Ions created by this process (H^+ or O^{2-}) diffuse through the electrolyte membrane to balance the reactions. At the same time, electrons flow from the anode to the cathode through an external circuit. If the fuel is hydrogen, the only waste product is water. Other fuels, such as methane, can be used, but they have to be converted to hydrogen before entering the cell.

Lately, evidence has emerged that methane can be converted directly by a solid-oxide fuel cell, without any flame, using cerium oxide as the electrode catalyst⁶. A typical solid-oxide fuel cell uses a hard ceramic material such as zirconia (which conducts O^{2-} ions) as the electrolyte, and has nickel-based anodes. Park *et al.*² now suggest that copper with cerium or samarium oxide may be a better anode catalyst for direct oxidation of more complex hydrocarbons that exist in realistic fuels such as kerosene or diesel. The challenge is to find pathways through which ethane and butane, or even aromatics such as toluene, can be reacted without fouling up the process through damaging side reactions that tend to produce tar or carbon.

This is the challenge that has been picked up by Park and colleagues. They used a copper catalyst intimately in contact with cerium oxide to produce a porous anode material. When this operated on pure butane for 48 hours there was no sign of carbon deposition, but good evidence of complete chemical conversion and electron transfer through a zirconia electrolyte membrane. An ordinary cell would die in minutes under the same operating conditions because of carbon fouling. When the copper–ceria–samarium catalyst was used, even toluene showed reasonable reactivity and electron conversion. This is very encouraging because aromatics are usually prone to rapid graphite deposition, which destroys the catalyst.

The hope is that even gasoline or diesel fuel may soon be catalytically converted to energy and heat in a clean and powerful manner by such devices. But it may be easier to use partly oxidized fuels such as methanol or formic acid because these will work in aqueous solutions with catalytic electrodes.

Such fuels react with oxidizing species at lower temperatures without carbon formation and may be ideal power supplies for computers or mobile phones⁷. But such intermediates, as with hydrogen or pure methane, might not be readily available in the quantities and economics we expect for regular fuels.

Of course, the fuel reactions in the device are not the only areas of ignorance. Fuel-cell systems are already semi-commercial but a number of issues remain to be resolved if the expected billion-dollar consumer market is to emerge in 2004 — the date set for car manufacturers in California to produce 5% zero-emission vehicles. One problem is the cost of electrolyte membranes, for both solid polymer and ceramic fuel cells, so the cost of materials will need to fall by an order of magnitude. The performance of a fuel cell may also be limited by peripheral equipment, which may represent 75% of the total device. Existing systems, such as the d.c./a.c. converter and heat exchanger, may not be well matched to the new fuel-cell technology. And particular

problems, such as leakage through the polymer membrane or cracking of the ceramic electrolyte, remain to be sorted out. It is worth recalling that Nernst, the inventor of the zirconia electrolyte in the 1890s, praised the simultaneous invention of the telephone because it enabled him to call home to switch the device on, the warm-up time being so long. Solutions to these problems are now being found⁸. Only when this happens will the expected applications in domestic heaters, hybrid vehicles and distributed power plants become reality. ■

Kevin Kendall is in the School of Chemical Engineering, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.
e-mail: k.kendall@bham.ac.uk

1. Grove Symposium *J. Power Sources* March (2000).
2. Park, S., Vohs, J. M. & Gorte, R. J. *Nature* **404**, 265–267 (2000).
3. Houghton, J. *Global Warming* Ch. 3 (Lion, Oxford, 1994).
4. Davy, H. *Phil. Trans. R. Soc. Lond.* **107**, 45–85 (1817).
5. Grove, W. R. *Phil. Mag.* **14**, 127–130 (1839).
6. Murray, E. P., Tsai, T. & Barnett, S. A. *Nature* **400**, 649–651 (1999).
7. Büchi, F. N. (ed.) *Portable Fuel Cells Conference* (Bossel, Switzerland, 1999).
8. Kendall, K. US Patent No. 5,827,620 (1998).

Evolution

In search of the whales' sisters

Zhexi Luo

Which mammals are the closest relatives to cetaceans (dolphins, porpoises and whales)? That is, which mammalian group is their 'sister taxon'? There is a wide gulf between the morphological and molecular evolutionary studies on the question, for they give conflicting answers. In a paper in *Systematic Biology*, O'Leary and Geisler¹ highlight the importance of the early divergent lineages, or extinct fossil taxonomic groups, in resolving this intriguing problem.

The descent of whales from land-dwelling mammals is a compelling example of evolution. It is documented by a rich fossil record of intermediate forms spanning the land–water transition; and by morphological and molecular lines of evidence, both of which testify that cetaceans have a close affinity to the artiodactyl mammals, or ungulates with even-toed feet and double-pulley ankle, such as today's cow, camel and hippopotamus. But two central issues have yet to be settled. One is the identity of the cetacean sister taxon. The other is whether the cetaceans are 'nested' within the group of artiodactyls as a branch on the family tree of the living artiodactyls — in other words, whether cetaceans represent some ancient artiodactyls that became adapted to aquatic life.

What O'Leary and Geisler have done is to evaluate all the available morphological and

molecular evidence^{1,2}. They show that both issues hinge on the long-extinct fossil groups that were split early from extant cetaceans. The topology, or branching sequence, of the family tree of living artiodactyls and cetaceans depends on inclusion of these divergent lineages of cetaceans and their putative extinct relatives.

Among those possible fossil relatives are the mesonychids that lived from 60 to 30 million years ago. Their feet were even-toed and adapted for running, like those of modern artiodactyls, and their fish-eating or carnian teeth closely resemble those of the earliest whales. Since the 1960s³, several morphological studies^{1,2,4,5} (with one exception⁶) have considered that extinct mesonychids and cetaceans are sister taxa (Fig. 1a). The combined mesonychid–cetacean group is, in turn, related to living artiodactyls. In this scheme, living artiodactyls are recognized as a clade^{1,2,5,7}, a genealogical group that includes all descendants of a common ancestor.

In contrast, molecular studies^{8–12} have generally concluded that hippopotamuses and cetaceans are more closely related to each other than either group is to other living artiodactyls (Fig. 1b). Consequently, the artiodactyls would be a grade that has similar evolutionary adaptations, but not a genealogical clade. Moreover, some molecular

NATURE

100 YEARS AGO

On reading the letter of Prof. Dexter on "Assaults and Drunkenness" (p. 365), I notice that there is one great fallacy in the argument. When a man is intoxicated and commits an assault, the result is entered in police reports as "assault", the more serious offence overshadowing the less. So that, in all probability, many of the cases of assault referred to in the statement were also drunkenness, but were not tabulated as such. The temperance is an important element; for its variations are probably the cause of the change of character of the offences recorded. The same quantity of alcohol will, as has often been noticed, have very different effects in the summer and in the winter. In hot weather alcohol has a stimulating influence; this is much less marked in the winter, and during this season the sedative effect is certainly more noticeable. Studying Prof. Dexter's curves in this light, and assuming the absence of any other fallacies, we may reasonably conclude that the number of those arrested in drunkenness or its results varies but little throughout the year. Probably the same people supply the cases of drunkenness in winter and assaults in summer.

From *Nature* 15 March 1900.

50 YEARS AGO

In 1921 birds described as tits were observed to prise open the wax-board tops of milk bottles on the doorsteps in Swaythling, near Stoneham, Southampton, and drink the milk. This is the first known record of an act which has now become a widespread habit in many parts of England and some parts of Wales, Scotland and Ireland, and which has to date been practised by at least eleven species of birds. The spread of the habit is interesting, because of the problems of behaviour involved. How far did the individual birds learn the habit from each other, or invent it for themselves? If most of them learnt it, by what process did they do so? How did, and how do, they detect the presence of food inside the bottle?... Without experimental evidence it is impossible to decide which senses are of use to the bird in indicating the presence of food. Several correspondents have found that bottles filled with water or even empty bottles are still attacked; but this conveys nothing if the previous history of the birds is not known.

From *Nature* 18 March 1950.

studies¹² suggest that mesonychids are not closely related to cetaceans; this is because the mesonychid–cetacean relationship implies a larger gap in the fossil record⁹, and is therefore less plausible than the hippo–whale connection¹². So, for all their congruence on the broad picture of ungulate–cetacean evolution, there is a big disagreement between morphological and molecular studies over these two phylogenetic issues.

What does that mean in terms of the origin of cetacean adaptations to aquatic life? If hippos and whales are sister taxa to the exclusion of mesonychids (as molecular studies

suggest), then cetacean adaptations such as underwater nursing of offspring and nearly hairless skin could have originated in the most recent common ancestor of both groups (Fig. 1b); this implies that certain aquatic features had evolved before the origin of cetaceans⁹. The similar ear regions^{4,5} and the fish-eating and carrion teeth must have evolved independently in mesonychids and in the earliest cetaceans because these derived features are absent in living hippos and their artiodactyl allies.

Alternatively, if mesonychids and cetaceans are sister taxa (as the best morphologi-

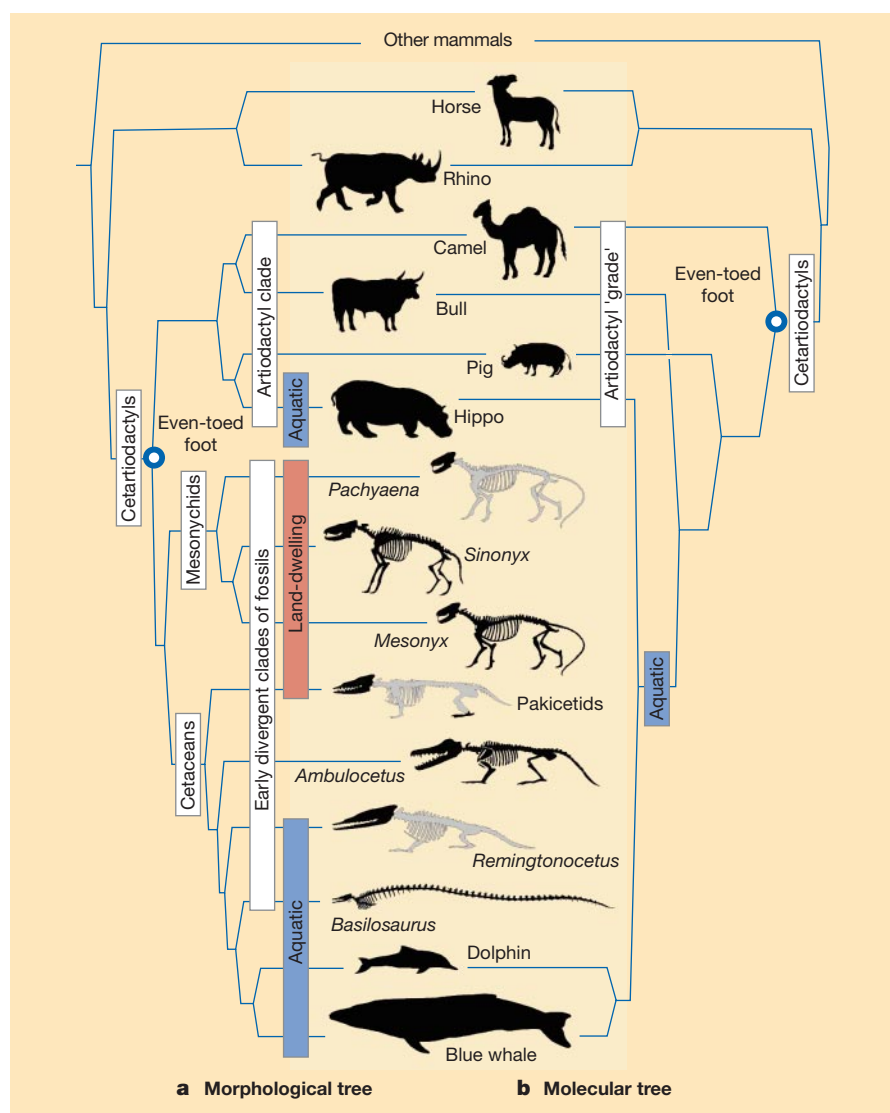


Figure 1 Different interpretations of cetacean evolutionary history. Central players in both phylogenetic trees are the even-toed ungulates, the artiodactyls (the other, odd-toed, group of ungulates includes animals such as the horse). a, A phylogenetic tree based on morphological evidence, including O'Leary and Geisler's extensive sampling of early divergent lineages¹. This scheme supports the traditional hypothesis that mesonychids are the sister taxon of cetaceans, and that artiodactyls are a monophyletic clade (a genealogical group that includes all descendants of a common ancestor) to the exclusion of living cetaceans. b, A tree based exclusively on molecular data from living ungulates and cetaceans. Here hippopotamuses and cetaceans are sister taxa, and artiodactyls are an evolutionary grade characterized by herbivory and even-toed feet for running. But they are not a good genealogical group unless cetaceans are considered to be a part of it. Living taxa are shown as silhouettes; early divergent taxa as skeletons; and incomplete fossils in grey. (Modified from refs 1, 2, 9 and 10.)

cal evidence has it), then the aquatic adaptations of hippos and living cetaceans must be convergences that occurred well after the split of their respective lineages. This is because certain primitive cetaceans (paki-cetids) have many ear⁴ and ankle⁶ structures typical of a land mammal, and mesonychids were fully terrestrial and adapted to running (Fig. 1a).

What about the relative strengths of the two lines of evidence? Morphological studies of ungulate–cetacean phylogeny take in a much wider range of taxa than the molecular studies. This is because the living artiodactyls and cetaceans available for molecular analyses represent only a few twigs of their bushy family trees that have survived pruning by extinction. As O'Leary and Geisler¹ point out, 90% of ungulate genera and more than 86% of the cetacean genera are extinct. Taking the mesonychid and cetacean fossil taxa into account produces a markedly different evolutionary history (Fig. 1a). From this, O'Leary and Geisler conclude that artiodactyls are a clade to the exclusion of cetaceans. Here we have a good example of the principle that including data from early divergent fossil lineages can shake, and reshape, the trees of living taxa based solely on molecular evidence.

However, for living taxa, use of sequences of genes and proteins is in some ways more powerful than use of morphological

characters. In recent years, more genes in larger sequence samples have been added to the arsenal for estimating ungulate and cetacean phylogeny^{9–12}. And the poor taxonomic sampling of some of the earlier molecular work has at least in part been redressed¹⁰. All in all, because molecular characters can vastly outnumber morphological features, they often prevail in simultaneous analyses of conflicting data sets. That is, in such analyses they effectively swamp the morphological estimates of ungulate–cetacean phylogeny.

Both morphological and molecular data are vulnerable to the problem of homoplasies — reversals to ancestral conditions or parallel changes in different lineages that can camouflage the true phylogeny. In this sense, neither approach is better than the other. For instance, the ear region of the skull, traditionally considered to be a good source of highly stable characters, shows some glaring homoplasies among the ungulates and cetaceans^{4,5}. Moreover, the fossil record of many early divergent fossil taxa is incomplete, resulting in ambiguities in morphological estimates.

On the molecular side, DNA and protein sequences have parallel and back mutations. Even the newest studies using retroposons, which are the RNA-mediated insertion sequences interspersed in the genome, have their limitations. Retroposons show a low

level of homoplasy^{12,13}, but mutational decay of the flanking region of retroposons may make them difficult to detect in older lineages. This means that retroposon-based estimates may not be effective for resolving lineages that go back more than 50 million years¹⁴. Cetaceans were already diversified by 53.5 million years ago, and their divergence from extant artiodactyls goes back much further^{15,16} than that.

A way to untie this Gordian knot may be to seek out compatible aspects of the molecular and morphological data sets. Measurement of the hidden support and conflict between them¹⁰ can help extract additional information. Also, morphological features should be better analysed for some living taxa for which extensive molecular data are available. Such studies may be just as helpful as discoveries of new fossils and genes in resolving details of the cetaceans' phylogenetic tree — most particularly, the question of which group is their sister taxon. ■

Zhexi Luo is in the Section of Vertebrate Paleontology, Carnegie Museum of Natural History, Pittsburgh, Pennsylvania 15213, USA.
e-mail: luoz@carnegiemuseums.org

1. O'Leary, M. A. & Geisler, J. H. *Syst. Biol.* **48**, 455–490 (1999).
2. O'Leary, M. A. *Cladistics* **15**, 315–330 (1999).
3. Van Valen, L. M. *Bull. Am. Mus. Nat. Hist.* **132**, 1–126 (1966).
4. Luo, Z. & Gingerich, P. D. *Papers on Paleontology (Univ. Michigan)* **31**, 1–98 (1999).
5. Geisler, J. H. & Luo, Z. in *Evolutionary Emergence of Whales* (ed. Thewissen, J. G. M.) 161–212 (Plenum, New York, 1998).

Fibre optics

Transparent talk

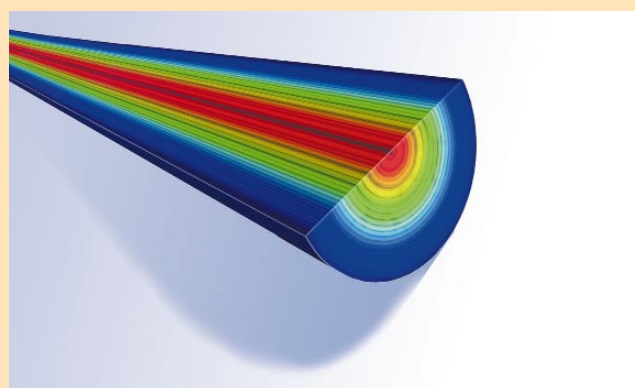
Modern telecommunications rely on the transmission of light signals along fibre-optic cables — fast, but how to minimize signal degradation over large distances? The secret is to eliminate contaminants that cause unwanted absorption. These intruders tend to creep into the fibres during the manufacturing stages. Elsewhere in this issue (*Nature* **404**, 262–264; 2000), Gordon Thomas and colleagues reveal how the main culprit, water, sneaks in. Putting this knowledge into practice allows the manufacture of dry fibres with near maximal transparency and improved bandwidth.

The problem with water, in particular the hydroxyl (OH) group, is that it gets excited when irradiated by certain infrared wavelengths. As this band is used for telecommunications, precious signal power can be lost in vibration of the water molecules.

So where does the water come from? One process for making

optical fibres involves heating silica rods with ultra-pure glass cores to very high temperatures. Once softened, the rods can be drawn out into fibres some tens of kilometres long. To achieve the necessary temperatures — in excess of 2,000 °C — it is common to use torches that burn hydrogen and oxygen: the perfect recipe for water.

But the incorporation of water into the silica rods is a more subtle matter. To investigate this process, Thomas and colleagues measured the transmission of infrared light through a small section of rod that had been cooled before drawing out, thereby 'freezing in' the water. They found that the absorption, and hence the hydroxyl concentration, is much stronger in the outer layers of the rod. This can be seen in the figure, which indicates the varying hydroxyl concentrations through a cross-section of the rod. (The colour scale is logarithmic, blue representing the highest concentrations.) Clearly, the water



diffuses from the outside in.

Given that the signals in optical fibres are confined to a narrow core region, the distribution of water in the rod might appear to be good news. But Thomas and colleagues found that the drawing process — which increases the aspect ratio by a factor of about a hundred million — lets the water in much further. They confirmed this by calculating the diffusion coefficient, a quantity that describes the flow of both water and glass during the contamination

process. This parameter was much higher than expected from low-temperature diffusion, probably reflecting an increased mobility of hydroxyl groups in the hot, molten state.

There is no doubt that absorption by hydroxyl groups contributes significantly to transparency loss in optical fibres. Uncovering the physical origin of the contamination points to an obvious solution: pick a water-free heat source.

Karen Southwell

6. Thewissen, J. G. M. & Madar, S. J. *Syst. Biol.* **48**, 21–30 (1999).
7. Luckett, W. P. & Hong, N. J. *Mamm. Evol.* **5**, 127–182 (1998).
8. Graur, D. & Higgins, D. *Mol. Biol. Evol.* **11**, 357–364 (1994).
9. Gatesy, J., Hayashi, C., Cronin, M. & Arcander, P. *Mol. Biol. Evol.* **13**, 954–963 (1996).
10. Gatesy, J., O'Grady, P. & Baker, R. H. *Cladistics* **15**, 271–313 (1999).
11. Ursing, B. M. & Arnason, U. *Proc. R. Soc. Lond. B* **265**, 2251–2255 (1998).
12. Nikaido, M., Rooney, A. & Okada, N. *Proc. Natl Acad. Sci. USA* **99**, 10261–10266 (1999).
13. Hillis, D. M. *Proc. Natl Acad. Sci. USA* **96**, 9979–9981 (1999).
14. Miyamoto, M. M. *Curr. Biol.* **9**, R816–R819 (1999).
15. Bajpai, S. & Gingerich, P. D. *Proc. Natl Acad. Sci. USA* **95**, 15464–15468 (1998).
16. Kumar, S. & Hedges, S. B. *Nature* **392**, 917–920 (1997).

Photonuclear physics

Laser light splits atom

Donald Umstadter

Lasers have become ubiquitous, being used in everything from a bar-code reader to a compact disk player. Who would have thought that they might be used to split the atom? A few scientists proposed to do just that more than a decade ago¹. But accomplishing it in the laboratory had to await the maturity of new technology², which enabled the construction of the world's most powerful lasers at the Lawrence Livermore National Laboratory in the United States and at the Rutherford Appleton Laboratory in the United Kingdom. Now two independent research teams have used these lasers to split the uranium atom^{3,4}. This work, reported in *Physical Review Letters*, is just the latest milestone in the race to discover what happens when matter interacts with the highest electromagnetic field strengths of light.

Over the past century, the peak power of artificial light sources has increased exponentially, from a kilowatt (10^3 watts) to a petawatt (10^{15} watts). This corresponds to an increase of more than a factor of ten each decade (Fig. 1). These dramatic advances have led to new physical insights and applications along the way. For instance, it was the invention of the cathode ray tube, used to make X-rays at the turn of the last century (and still used to this day for most television displays), that led to the discovery of electrons. Later, the invention of the laser increased the power of light to the point that it could accelerate an electron to close to the speed it naturally has as it orbits the nucleus of the atom. Remarkably, the distance required to reach this speed is but a millionth of a metre, the light's own wavelength. Having laser fields almost comparable to atomic fields has led to the discipline of physics known as nonlinear optics.

With the latest increase in laser power, electrons can now be accelerated to even higher speeds, close to the speed of light itself, again in a distance of just the laser wavelength. At such high speeds, electrons are governed by the laws of special relativity, leading to new types of nonlinear motion⁵. Gamma rays (energetic X-rays) are emitted when the highly energetic 'relativistic' elec-

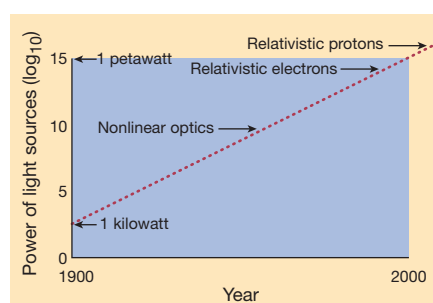


Figure 1 As the power of laboratory light sources increases, the energy of accelerated particles increases accordingly, making laser-induced nuclear reactions possible. The most powerful lasers have now reached petawatt (10^{15} watts) peak powers, allowing them to accelerate electrons to relativistic speeds. The gamma rays produced by such electrons have been used in recent experiments to split the uranium atom^{3,4}.

trons accelerate rapidly during a collision with neighbouring ions. If these gamma rays are energetic enough, they can then perturb the delicate stability of heavy nuclei such as uranium, causing it to split into lighter elements and neutrons in a process called photonuclear fission. This is exactly what the two groups of laser scientists have just demonstrated.

The fission of uranium usually brings to mind nuclear reactors. But before you get your hopes up that splitting atoms with lasers will solve the world's energy problems, it should be pointed out that more energy is required to power the laser than is released through the fission process. Besides, there are much more efficient ways to accelerate electrons and produce gamma rays than by using optical light — for instance, with accelerators driven by radio-frequency waves.

Because photofission has been studied with radio-frequency accelerators since they were invented decades ago, it is now quite well understood. In fact, a simple search of a database of physics journals for articles published since 1967 using the keyword 'photofission' yields a total of 463 hits. Only seven of these appeared in the journal *Physical Review Letters*. Usually, the number

of articles published in this exclusive physics journal on a particular topic tapers off as our fundamental knowledge of the subject increases. So why were references 1, 3 and 4 the only ones in the past 15 years to enjoy this privilege? As the authors readily admit, it seems less to do with what they tell us about photofission than with the tantalizing prospect that lasers can now be used to induce nuclear processes and the myriad applications that may follow.

The most familiar laser-driven nuclear process is fusion (the inverse of fission), in which lasers are used to heat hydrogen nuclei to temperatures greater than that of the Sun in order to fuse them together to make helium and release energy. Indeed, one of the main reasons for the current interest in the fission papers^{3,4} is that the relativistic electrons produced by these experiments might some day be used as a spark plug to ignite a fusion reaction.

Another laser-driven nuclear reaction that might soon be possible is the creation of an ionized gas (plasma) of positrons as well as electrons. (Positrons are anti-electrons.) High-density antimatter is difficult to produce on Earth because it annihilates upon encountering ordinary matter, but it is of interest to astrophysicists because it might exist elsewhere in the Universe, for example in pulsars, quasars and black holes. Remarkably, with compact powerful lasers, dense electron-positron plasmas might soon be available for study in university laboratories.

Several groups have used electrons accelerated by high-power lasers to accelerate ions to high velocities^{6,7}. The ions can be used to produce high concentrations of short-lived radioactive isotopes to tag cancer cells in nuclear medicine. More energetic protons from larger and more expensive radio-frequency-driven accelerators are even being used to kill cancer cells.

As the power of lasers continues to increase, the day will come when protons can even be accelerated to relativistic velocities and, possibly, directly by the laser field itself. (Protons are more difficult to accelerate than electrons because they are thousands of times more massive.) This next era of laser-matter interactions will involve all sorts of other nuclear reactions that could lead to novel applications, some of which are anyone's guess.

Donald Umstadter is at the Center for Ultrafast Optical Science and the Department of Nuclear Engineering and Radiological Science, University of Michigan, Ann Arbor, Michigan 48109, USA. e-mail: dpu@umich.edu

1. Boyer, K., Luk, T. S. & Rhodes, C. K. *Phys. Rev. Lett.* **60**, 557–559 (1988).
2. Strickland, D. & Mourou, G. *Opt. Commun.* **56**, 219–221 (1985).
3. Ledingham, K. W. D. *et al. Phys. Rev. Lett.* **84**, 899–902 (2000).
4. Cowan, T. E. *et al. Phys. Rev. Lett.* **84**, 903–906 (2000).
5. Chen, S.-Y., Maksimchuk, A. & Umstadter, D. *Nature* **396**, 653–655 (1998).
6. Sarkisov, G. *et al. Phys. Rev. E* **59**, 7042–7054 (1999).
7. Krushelnick, K. *et al. Phys. Rev. Lett.* **83**, 737–740 (1999).

The element of uncertainty

Russell J. Hemley

Consider a box containing an equal number of electrons and protons — what could be simpler? If the box is large enough, this ensemble of particles spontaneously forms a collection of hydrogen molecules (H_2), as both theory and experiment most certainly agree. But as the size of the box is reduced, the system evolves into a fully quantum-mechanical, many-body problem that defies an exact theoretical solution, because approximations that work in the low-density case break down. Yet precisely what happens to this 'simple' system when the box is compressed is a key problem in physics, chemistry, planetary science and astrophysics. A report by Kitamura *et al.*¹ on page 259 of this issue, together with two other theoretical studies published in *Nature*² and *Physical Review Letters*³, tackle this problem using different approaches. The results show signs of an emerging consensus but at the same time a diversity of predictions that lead to new puzzles.

Solid hydrogen was created in the laboratory just over a century ago, and has been studied under a growing range of conditions in recent years. It was first suggested in 1935 that hydrogen molecules should break up at sufficiently high pressures to form a monatomic metallic phase⁴. Or, from a chemical viewpoint, its halogen-like (diatomic) character gives way to that of an alkali metal. Wigner and Huntington⁴ calculated that this would occur above 25 gigapascals (25 GPa, or 250,000 times atmospheric pressure). But as experiments carried out over the past decade have shown, the picture is not so simple. The covalent bond between hydrogen pairs is particularly tenacious and persists to pressures of more than 300 GPa, and pressure-induced phase transitions occur within the molecular state whose character, and even existence, was not predicted theoretically⁵. Understanding these experimental observations from first principles is the challenge faced by this new set of calculations.

The hydrogen molecule is a system of elegant symmetry but only putative simplicity: with no inner core of electrons, its physical properties ought to be derived simply from point charge interactions between the protons and electrons. But for an interacting ensemble of particles, most calculations have to rely on a hierarchy of approximations to the Schrödinger equation — the quantum-mechanical description of the system. One of the most common approximations is the treatment of the nuclei as classical objects. As dictated by the uncertainty principle, however, the low mass of the proton makes

the nuclei quantum-mechanical objects. This gives rise to a substantial zero-point motion when the protons are 'at rest' and the persistence of essentially free rotation of the molecules at megabar pressures (~ 100 GPa). So, unlike other molecular systems, a complete theoretical solution requires that

the nuclei and electrons are treated on the same quantum-mechanical footing.

Kitamura and colleagues' simulations¹ take into account the quantum character of the protons. With no guidance from experiments, the system spontaneously forms three different structures under different pressures and temperatures, corresponding to the three molecular phases of dense hydrogen found experimentally (Fig. 1a). These are a high-temperature, low-pressure phase (phase I), a low-temperature, high-pressure phase (phase II) and a higher-pressure phase (phase III). The results confirm earlier work

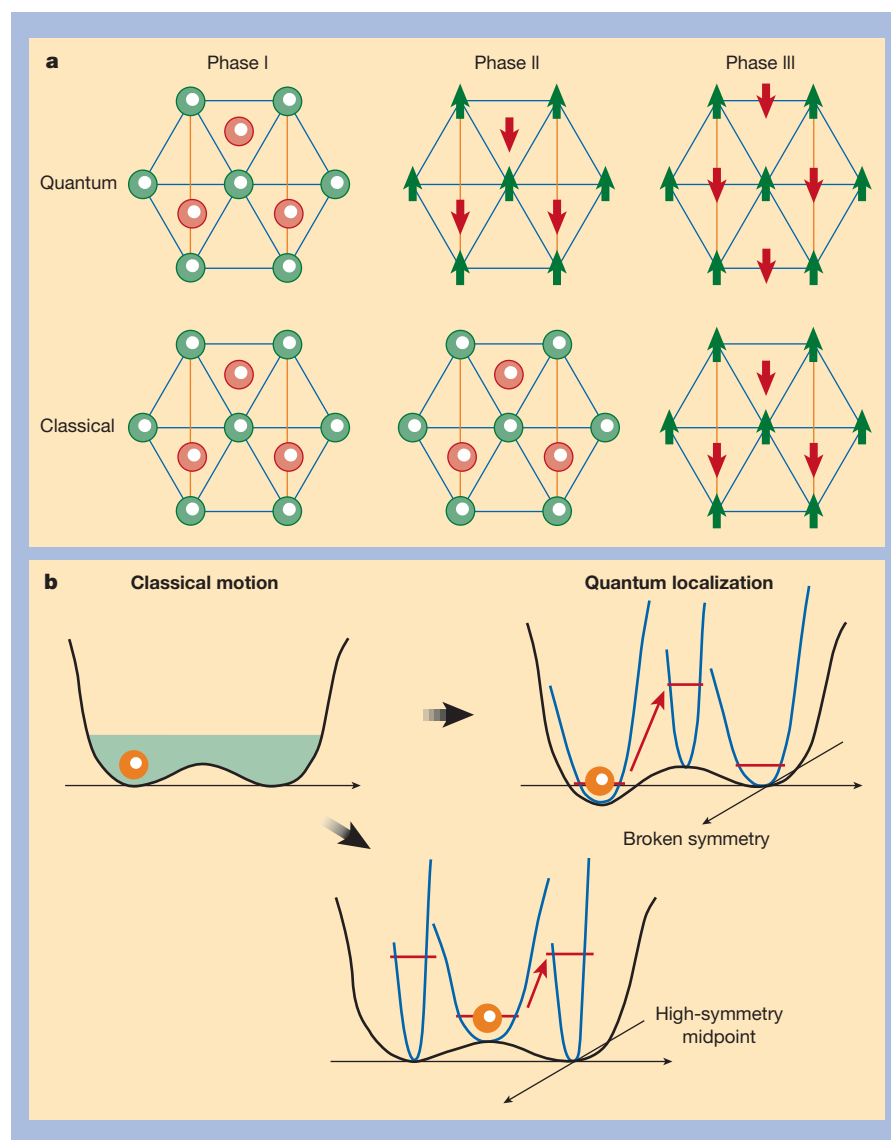


Figure 1 Solid molecular hydrogen at high pressures. **a**, The proton distributions obtained by quantum and classical simulations for phases I, II and III of dense H_2 . The arrows show that the molecules are orientated in a particular direction, whereas spheres indicate that the molecules are rotating. Red and green denote the first and second layers, respectively. The quantum-mechanical simulations of Kitamura *et al.*¹ reproduce aspects of the ordering observed experimentally for phases II and III. **b**, In a classical simulation, the particle moves unquantized on the potential energy surface (which may be one-, two- or multidimensional). In the quantum simulation, the particle can tunnel through the barriers leading to quantum localization. Depending on the shape of the multidimensional potential surface, the particle can be localized on one side or the other to give a phase of broken symmetry (as predicted for phase II of hydrogen), or localized at a high-symmetry midpoint, as predicted for muonium in silicon (ref. 1; courtesy of S. Tsuneyuki).

showing that rotational order in phase III is largely classical, whereas that in phase II is purely quantum mechanical⁵. Moreover, the calculations reveal an unexpected quantum localization of the proton density in phase II (Fig. 1b). Solid hydrogen seems to reflect the same physics predicted for another quantum system — dense ice — which showed the quantum character of the protons evolving towards a classical symmetric hydrogen-bonded state at higher pressure⁶.

A different tack is taken by Johnson and Ashcroft² in their *Nature* paper, which addresses the question of band overlap in hydrogen, the prediction that at a critical pressure some bonding electrons move into the conduction band to produce a molecular metallic solid⁷. The predicted pressure at which this transition occurs is also a source of uncertainty, in part because of limitations in the approximation to the Schrödinger equation used to study electron bands — the 'local density approximation'. Moreover, the predicted transition depends strongly on crystal structure and dynamics; there are many ways in which the hydrogen molecules can orientate themselves to prevent the loss of bonding electrons to the conduction band.

Johnson and Ashcroft² use a simple model that corrects shortcomings of the local density approximation arising from the neglect of many-electron effects. The simplicity of this approach enables them to examine a large number of structures and to estimate the important effects of proton dynamics. Their analysis predicts that an insulating molecular state will persist to ~400 GPa, at which point band overlap occurs and the ever-tenacious covalent bond endures to still higher pressures in a molecular metallic state.

Different findings emerge from the related calculations of Kohanoff *et al.*³ published in *Physical Review Letters*. They simulate the high-pressure phases of H₂ in an effort to reproduce in detail the observations from vibrational spectroscopy experiments, which provide indirect data on structure and symmetry. There is a sudden onset of strong infrared absorption of the hydrogen 'stretching mode' in phase III (ref. 8), which is notable for its absence in the isolated molecule. This observation indicates a symmetry-breaking charge transfer that has been the touchstone, if not the bane, of theoretical models of hydrogen^{8–12}.

Kohanoff *et al.*³ treat this charge transfer using a quadrupole-induced dipole model⁵, and go on to reproduce rich features of the low-frequency spectrum that suggest a vibrational instability, perhaps related to a spontaneous polarization calculated previously^{10,11}. But the most important result of this work is the prediction of yet another class of high-pressure structures, one based on a cluster of three H₂ molecules. It is reminiscent of a supramolecular complex proposed

previously⁹, and going back even further, to the predicted stability at high pressure of a 'benzene-like' H₆ complex obtained in quantum-chemical calculations¹³.

The story of this 'simple system' of electrons and protons is far from complete. The original study of Wigner and Huntington⁴ hinted at the underlying complexity of the system by suggesting, from chemical arguments, that layer-like structures might be seen before the formation of the monatomic metallic phase. If intermediate molecular metallic states exist, they may exhibit novel properties such as very high-temperature superconductivity¹⁴. The latest set of theoretical results should be tested by the broad range of diffraction, magnetic and electrical techniques being developed for studies at multimegabar pressures (> 300 GPa). The goal is a complete understanding of the system, including the origin of the electrical conductivity observed in the hot, shock-compressed fluid¹⁵. Experience indicates that addressing these issues may lead to the unexpected: if hydrogen is the

element of uncertainty, it is also the element of surprise. ■

Russell J. Hemley is at the Geophysical Laboratory and Center for High-Pressure Research, Carnegie Institution of Washington, 5251 Broad Branch Road NW, Washington DC 20015-1305, USA.

e-mail: hemley@gl.ciw.edu

1. Kitamura, H., Tsuneyuki, S., Ogitsu, T. & Miyake, T. *Nature* **404**, 259–262 (2000).
2. Johnson, K. A. & Ashcroft, N. W. *Nature* **403**, 632–635 (2000).
3. Kohanoff, J., Scandolo, S., de Gironcoli, S. & Tossati, E. *Phys. Rev. Lett.* **83**, 4097–4100 (1999).
4. Wigner, E. & Huntington, H. B. *J. Chem. Phys.* **3**, 764–770 (1935).
5. Mazin, I. I., Hemley, R. J., Goncharov, A. F., Hanfland, M. & Mao, H. K. *Phys. Rev. Lett.* **78**, 1066–1069 (1997).
6. Benoit, M., Marx, D. & Parrinello, M. *Nature* **392**, 258–261 (1998).
7. Friedli, C. & Ashcroft, N. W. *Phys. Rev. B* **16**, 662–672 (1977).
8. Hemley, R. J., Soos, Z., Hanfland, M. & Mao, H. K. *Nature* **369**, 384–387 (1994).
9. Tse, J. S. & Klug, D. D. *Nature* **378**, 595–597 (1995).
10. Edwards, B. & Ashcroft, N. W. *Nature* **388**, 652–655 (1997).
11. Souza, I. & Martin, R. M. *Phys. Rev. Lett.* **81**, 4452–4455 (1998).
12. Nagao, K., Takezawa, T. & Nagara, H. *Phys. Rev. B* **59**, 13741–13753 (1999).
13. Dixon, D. A., Stevens, R. M. & Herschbach, D. R. *Faraday Discuss. Chem. Soc.* **62**, 110–126 (1977).
14. Richardson, C. F. & Ashcroft, N. W. *Phys. Rev. Lett.* **78**, 118–121 (1997).
15. Weir, S. T., Mitchell, A. C. & Nellis, W. J. *Phys. Rev. Lett.* **76**, 1860–1863 (1996).

Cell biology

Reeling CASK into the nucleus

David S. Bredt

When cells touch one another as they grow and migrate, they erect specialized structures at the sites of contact. These 'junctional complexes' form bridges between adjacent cells and link cell-surface adhesion molecules to the underlying framework of the cell, the cytoskeleton. The complexes also serve as communication centres that relay signals from neighbouring cells. Signals from cell junctions are often carried by cascades of signalling molecules to the nucleus, where they regulate cell growth and differentiation. In contrast to this classical pathway, Hsueh *et al.* report on page 298 of this issue¹ a new signalling pathway in which CASK, a structural girder of cell junctions², directly enters the nucleus and thereby regulates gene expression.

CASK is a member of the membrane-associated guanylate kinase (MAGUK) family, a group of conserved cytoskeletal proteins that are composed of arrayed modular domains. MAGUKs all contain characteristic areas called PDZ and SH3 motifs, which mediate specific protein–protein interactions and form scaffolds for protein networks at cell membranes³. For example, the PDZ domain of CASK binds specifically to a transmembrane adhesion protein called syndecan, concentrating it at epithelial cell junctions⁴. Many proteins contain PDZ and SH3 domains, but the signature for MAGUKs is a carboxy-terminal guanylate

kinase-like (GUK) motif. CASK also contains a second kinase-like domain at the amino terminus. The functions of these two domains are mysterious, as neither appears to have enzymatic activity.

Hsueh *et al.*¹ screened for proteins that would bind to the GUK domain of CASK. Surprisingly, they found that the GUK domain associates with T-brain-1 (Tbr-1), a T-box transcription factor that normally resides in the nucleus⁵. The prototypical member of the T-box family is Brachyury, which is mutated in a strain of short-tailed (T) mice — hence the name 'T-box'. All proteins in the Brachyury family contain a T-box motif, which binds to a specific DNA element⁶. Interestingly, CASK associates with a unique carboxy-terminal domain of Tbr-1 outside the T-box and does not interact with Brachyury.

At first sight, the connection between CASK and Tbr-1 seems perplexing, as these proteins reside in distinct parts of the cell — the cytoskeleton near the cell membrane and the nucleus, respectively. A solution to this paradox is provided by Hsueh and colleagues' experiments¹ showing that, in cells that have been transfected with both CASK and Tbr-1, CASK relocates from the cytoplasm to the nucleus. Remarkably, this recruitment of CASK to the nucleus enhances Tbr-1-dependent transcription of reporter genes. This transcriptional

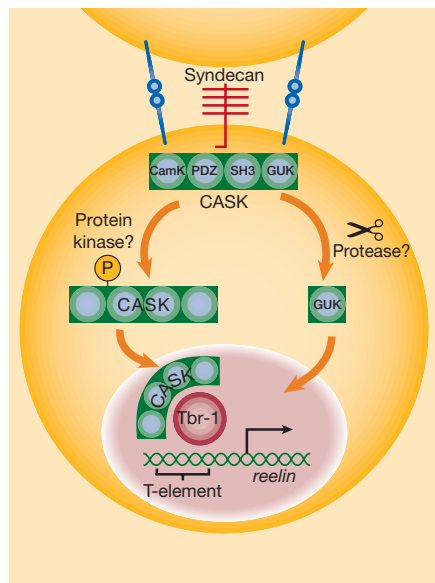


Figure 1 Nuclear translocation of CASK and coactivation of Tbr-1-dependent reelin expression. CASK is a component of the cortical cytoskeleton, and its PDZ domain binds to the cytosolic tail of syndecan, a cell-surface adhesion molecule. Hsueh *et al.*¹ have shown that signalling at the cell surface, perhaps transduced by a protein kinase or protease, dislodges CASK from the cytoskeleton. CASK, or its isolated GUK domain, then translocates into the nucleus in association with the Tbr-1 transcription factor and activates expression of reelin by binding to the T-element consensus sequence.

coactivation requires only the GUK domain of CASK and is independent of the T-box domain of Tbr-1.

Although Brachyury and Tbr-1 each bind a sequence of DNA known as the 'T-element' consensus site, the specific genes regulated by T-box transcription factors have remained elusive. One proposed target of Tbr-1 is the *reelin* gene, as both Tbr-1 knockout mice⁵ and *reeler* mutant mice⁷ exhibit abnormal cell layering in the developing brain. Furthermore, the *reelin* promoter region contains a T-element for Tbr-1 binding. To test for regulation of reelin by CASK, Hsueh *et al.* transfected neurons with a *reelin* reporter together with CASK or Tbr-1 (or both). Activation of this reporter was greatly increased by CASK and was further augmented by the combination of CASK and Tbr-1, so reelin could be a physiological target for the Tbr-1/CASK complex¹. However, regulation of reelin expression by Tbr-1 or CASK *in vivo* has not yet been established.

A particularly striking aspect of this story is the migration of CASK, a cytoskeletal element enriched at cell junctions, to the nucleus. By their design, cell adhesions are highly stable structures — indeed, junctional protein complexes are often provisionally defined by their cement-like attachment to the cytoskeleton beneath the cell membrane. Nevertheless, there is a precedent.

Another cytoskeletal protein, β -catenin, also migrates to the nucleus. Extracellular signalling by the *Wnt* oncogene triggers a cascade of cellular protein phosphorylation that causes β -catenin to accumulate in the cytoplasm. It then binds to Tcf/LEF transcription factors, forming a complex that translocates to the nucleus to regulate gene expression⁸.

So, what triggers CASK to migrate to the nucleus? By analogy to β -catenin, it is possible that protein phosphorylation may regulate translocation of CASK (Fig. 1). Possible kinases that might trigger CASK movement include protein kinase C, which binds to syndecan-4 (ref. 8), and the brain-derived neurotrophic factor (BDNF)-stimulated tyrosine kinase, which regulates the expression of reelin⁹.

Alternatively, translocation of CASK may be regulated by a relative balance of protein interactions, as Hsueh *et al.* found that overexpression of syndecan can regulate nuclear accumulation of CASK. Finally, evidence that the isolated GUK domain of CASK can coactivate Tbr-1 hints at proteolytic processing of CASK and release of the GUK domain to regulate nuclear translocation and gene expression (Fig. 1). There is precedent for such a mechanism: cleavage of the transmembrane protein Notch frees its carboxy-terminal tail to translocate to the nucleus. Proteolytic processing of Notch requires presenilin-1 (PS1), which is mutated in familial Alzheimer's disease¹⁰. Mice lacking PS1 have structural brain abnormalities that resemble those of Tbr-1 and *reeler* mutant mice, suggesting that PS1 could regulate translocation of CASK.

Hsueh and colleagues' results provide a new mechanism whereby signals from cell junctions can regulate gene expression in neurons. Whereas Tbr-1 expression is largely restricted to developing brain, CASK and other MAGUK proteins, such as the postsynaptic density protein PSD-95, are expressed most abundantly at synaptic junctions in mature neurons. It will be particularly interesting to see whether neuronal activity can induce nuclear translocation of these synaptic MAGUKs, and whether this participates in aspects of synaptic plasticity. ■

David S. Bredt is in the Department of Physiology, University of California at San Francisco School of Medicine, 513 Parnassus Avenue, San Francisco, California 94143-0444, USA.
e-mail: bredt@itsa.ucsf.edu

1. Hsueh, Y.-P., Want, T.-F., Yang, F.-C. & Sheng, M. *Nature* **404**, 298–302 (2000).
2. Hata, Y., Butz, S. & Sudhof, T. C. *J. Neurosci.* **16**, 2488–2494 (1996).
3. Craven, S. E. & Bredt, D. S. *Cell* **93**, 495–498 (1998).
4. Cohen, A. R. *et al.* *J. Cell Biol.* **142**, 129–138 (1998).
5. Bulfone, A. *et al.* *Neuron* **21**, 1273–1282 (1998).
6. Kispert, A., Koschorz, B. & Herrmann, B. G. *EMBO J.* **14**, 4763–4772 (1995).
7. D'Arcangelo, G. *et al.* *Nature* **374**, 719–723 (1995).
8. Nusse, R. *Cell* **89**, 321–323 (1997).
9. Ringstedt, T. *et al.* *Neuron* **21**, 305–315 (1998).
10. Haass, C. & De Strooper, B. *Science* **286**, 916–919 (1999).

Daedalus

Electric flight

When a gas molecule hits a solid surface, it bonds for a short while before flying off. Suppose that surface is carrying an electric current. Its outer electrons, responsible both for bonding and for carrying current, will be moving. So when the molecule disengages, it will have acquired the velocity of the current.

For normal currents, this velocity will be very low compared to its random thermal velocity. But, says Daedalus, in a thin-film superconductor, the current density might be made so high that the electrons reach velocities of many metres per second. A gas molecule hitting such a surface would be fired off with a significant added velocity in the direction of the current. This electro-streaming effect would exert thrust in the opposite direction. The current would feel 'aerodynamic resistance' as energy was extracted from it to propel the adjacent gas molecules. The surface would become an elegant, efficient, silent aero-engine with no moving parts.

So when a room-temperature superconductor finally emerges from some solid-state research laboratory, Daedalus will be ready to apply it. His range of 'electrofoil' flying machines will generate thrust over their entire surface. In effect they will be 'gliders with gain'. They will have amazing manoeuvrability and lift, as they will be controlled and steered by varying the surface current at all points, not just at a few control surfaces. And an electrofoil wing, by reducing or even reversing the current on its underside, could induce a Magnus vortex giving far more lift than that of any passive wing.

Electrofoil aircraft will probably not be very large or fast. Even with magnets in their wings, to cancel the field due to the current, their superconducting current density will be limited. But they will make ideal small aircraft: probes, pilotless vehicles, high-altitude relay planes, and so on. Even the snoopers' dream, a tiny, silent, insect-like plane with a TV camera in its nose, that can be steered remotely anywhere, even inside the most secure building, will be feasible at last. Meanwhile, Daedalus is testing his ideas with the best cryogenic superconductors. They should be able to create thrust in cold neon gas flowing over them, or pump liquid hydrogen for rocket motors. **David Jones**

The Further Inventions of Daedalus (Oxford University Press), 148 past Daedalus columns expanded and illustrated, is now on sale. Special *Nature* offer: m.curtis@nature.com

Growth of carbon micro-trees

Carbon deposition under extreme conditions causes tree-like structures to spring up.

The complex deposition and growth of graphitic carbon has come up with several surprises in recent years, including the growth of fullerenes and carbon nanotubes¹. Here we describe the spectacular growth of micron-sized tree-like carbon structures produced by chemical vapour deposition of methane without the use of any catalysts. Our results show that carbon surfaces can be tailored while they are growing simply by controlling the rate of deposition from the vapour phase.

The remarkable growth of carbon trees occurs autocatalytically, through the restructuring of newly deposited carbon surfaces. Figure 1 shows an overview of carbon trees of varying heights formed on graphite electrodes. The trees taper outwards from the surface and the stalks have the shape of narrow inverted cones. Growth proceeds during many intermittent cycles in what we call a 'flash CVD' process.

Flash CVD involves fast resistive heating of graphite electrodes in an atmosphere of methane and helium at a total pressure of 500 torr. The temperature of the heated zone on the electrodes, where the trees grow, ranges between 1,100 and 2,200 °C. During each cycle (lasting about 30 seconds), heating occurs extremely rapidly at rates of several hundred degrees centigrade per second. Trees nucleate during different cycles, but growth continues at all cycles at an average rate of 200 $\mu\text{m min}^{-1}$ (Fig. 1). The diameter of the smallest tree we observed was about 100 nm.

What stimulates this unusual growth of carbon structures from plain graphite surfaces? During the CVD flashes, carbon deposits from the vapour phase as a new graphitic surface with a strong spherulitic (pyrolytic graphite structure) morphology². During growth, several spherulitic nodules that are almost spherical in shape (Fig. 2a) emerge out of the surface. An explanation

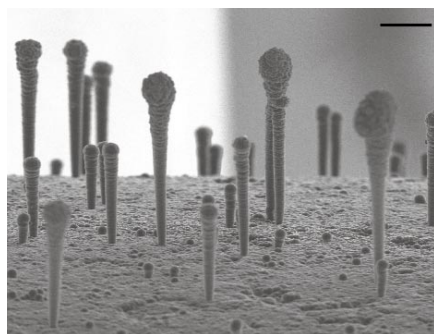


Figure 1 Carbon trees of different heights form on graphite electrodes, growing autocatalytically through the restructuring of newly deposited carbon surfaces. Scale bar, 100 μm .

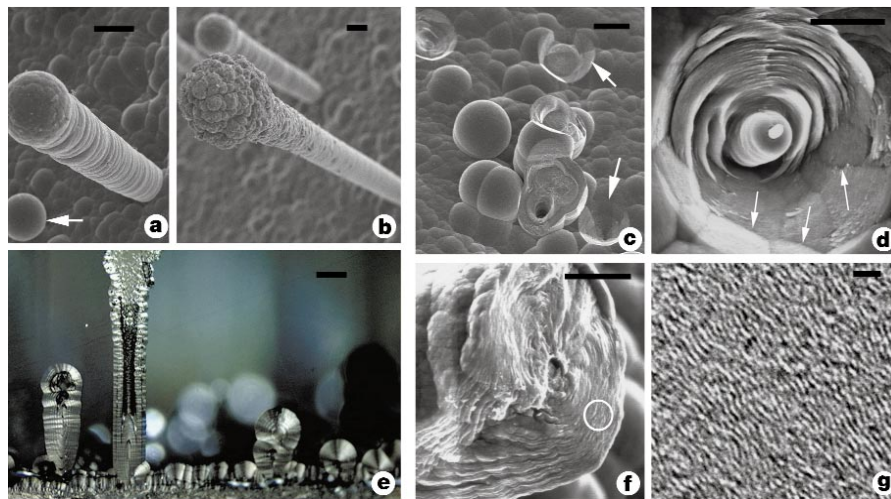


Figure 2 Morphology and structure of autocatalytically grown carbon micro-trees. **a–d,f**, Scanning electron microscope images of carbon micro-trees. **a**, Short trees with smooth heads; **b**, taller trees with rough heads; **c**, craters (arrows) that broken trees have left behind on the surface; **d**, a crater with the inner tree core still attached (arrows indicate lines that divide adjacent grains grown together during deposition); **f**, stem of a broken carbon micro-tree. **e**, Optical micrograph of trees embedded in epoxy resin and polished to reveal cross-sections. **g**, High-resolution transmission electron microscope image from the outer region (circle in **f**) of a tree, showing disordered graphite layers. Scale bars: 10 μm in **a–f**, and 1 nm in **g**.

for this is the fact that the nodules (made of turbostratic graphite layers) are misoriented with respect to each other, and so the rotation of planes at the nodule–nodule interfaces is restricted. This limits the planarization of the surface, hence these nodules serve as the nuclei for subsequent outward growth (Fig. 2a). The trees increase in height with successive cycles, until, above a certain critical height, the ends break down into a finer morphology (Fig. 2b). Polished cross-sections of the trees, observed by optical microscopy, reveal that the trees are hollow, especially near the base (Fig. 2e).

Trees broken off from the base give rise to conical craters on the fracture surfaces. These craters show a clear delineation (Fig. 2c) of individual grains that runs normal to the original surface, indicating that adjacent spherulites grew simultaneously upwards during deposition. The trees could have originated from grains that grew at abnormally high growth rates. The structure of the broken trees and the morphology of the craters suggest that there are at least two zones present in the microstructure (Fig. 2d). The outer regions show layering, with layers normal to the tree axis (Fig. 2f). The inner core is solid and there is often empty space between distinct zones.

To understand the atomic structure, we made cross-sectional specimens from individual trees and studied them by transmission electron microscopy (TEM). Samples

were prepared by ultramicrotoming trees embedded in polymer resins, as well as by slicing trees placed on substrates, and were observed by using a focused ion beam microscope. High-resolution TEM images of the outer zones show the structure of disordered graphite (Fig. 2g), with an interlayer distance of 0.342 nm, and confirm the general orientation of layers seen in Fig. 2f. The solid-like centre zone shows more disorder (in electron diffraction), indicating that the cores experience lower growth temperatures.

We have shown that the deposition of carbon under extreme conditions (that is, using rapid heating and cooling cycles) can generate structures with very unusual morphologies. This is unlike most CVD processes, where carbon structures grow with the help of catalyst particles³.

**P. M. Ajayan*, J. M. Nugent*,
R. W. Siegel*, B. Wei†,
Ph. Kohler-Redlich†**

**Department of Materials Science and Engineering, Rensselaer Polytechnic Institute, Troy, New York 12180-3590, USA*

†Max-Planck-Institut für Metallforschung, Seestr. 92, 70174 Stuttgart, Germany

1. Dresselhaus, M. S., Dresselhaus, G. & Eklund, P. C. *Science of Fullerenes and Carbon Nanotubes* (Academic, New York, 1996).

2. Harvey, J., Clark, D. & Eastbrook, J. N. in *Special Ceramics* (ed. Popper, P.) 181–202 (Proc. Symp. Brit. Cer. Res. Assoc., Academic, London, 1963).

3. Endo, M., Saito, R., Dresselhaus, M. S. & Dresselhaus, G. in *Carbon Nanotubes* (ed. Ebbsen, T. W.) 35–111 (CRC Press, Boca Raton, 1997).

Human performance

Scaling in athletic world records

World records in athletics provide a measure of physical¹ as well as physiological human performance^{2,3}. Here we analyse running records and show that the mean speed as a function of race time can be described by two scaling laws that have a breakpoint at about 150–170 seconds (corresponding to the ~1,000 m race). We interpret this as being the transition time between anaerobic and aerobic energy expenditure by athletes.

Records measured under standard external conditions represent the most reliable and up-to-date index of human performance, and attempts have been made to model running records since the beginning of the century^{1,4,5}. It has been suggested¹ that the record times, τ , and distances, d , of men's running can be represented by a single power law, $\tau \approx d^n$, in races from 100 to 10,000 metres, where n is about 1.1 and depends on the epoch. This power law implies that there is an invariance of statistical features of the system over all distances and there are no evident characteristic scales as in critical phenomena.

Table 1 Scaling exponents and break-time for world running records

Year	Men			Women		
	β_{an}	β_{ae}	τ^* (min)	β_{an}	β_{ae}	τ^* (min)
1919	0.175 ± 0.011	0.073 ± 0.007	3.52	-	-	-
1929	0.168 ± 0.007	0.072 ± 0.005	3.87	-	-	-
1939	0.163 ± 0.014	0.077 ± 0.003	3.91	-	-	-
1949	0.159 ± 0.013	0.079 ± 0.003	3.45	-	-	-
1959	0.157 ± 0.012	0.077 ± 0.004	3.33	-	-	-
1969	0.169 ± 0.012	0.069 ± 0.003	2.96	-	-	-
1979	0.161 ± 0.008	0.070 ± 0.002	3.22	0.175 ± 0.010	0.079 ± 0.007	3.42
1989	0.158 ± 0.011	0.071 ± 0.002	3.21	0.181 ± 0.014	0.069 ± 0.004	3.03
1999	0.165 ± 0.008	0.072 ± 0.003	2.55	0.175 ± 0.008	0.071 ± 0.004	2.83

Scaling exponents β_{an} and β_{ae} and time at the break, τ^* , are shown for running records at different epochs (marathons have been included from 1959).

We find, however, that both running and swimming can be characterized by two distinct critical phenomena when the record mean speed, u , which is equal to d/τ , is considered. Unlike d , which is set, u relates to an athlete's energy expenditure. The exponent of $\tau \approx d^n$ is almost 1, so the mean speed is $u = d/\tau \approx \tau^{-\beta}$, where the exponent $\beta = 1 - 1/n$. The fact that n is slightly larger than 1 amplifies every small deviation in the plot of d against τ in a plot of u against τ .

Using record data for swimming and running for men and women, we find that there are two regimes described by two scaling laws of the kind $u \approx \tau^{-\beta}$, where β has two different values. These separate abruptly at a characteristic time, τ^* , which lies in

the range 2.2–2.8 min (Fig. 1) that separates short races ($200 \text{ m} \leq d \leq 1,000 \text{ m}$ and $50 \text{ m} \leq d \leq 200 \text{ m}$ for running and swimming, respectively) and long races ($1,500 \text{ m} \leq d \leq 42,195 \text{ m}$ and $400 \text{ m} \leq d \leq 1,500 \text{ m}$ for running and swimming, respectively). Two different critical phenomena (and no more) simultaneously coexist in speed sports.

The transition between the two β scaling exponents corresponds to the switch from the anaerobic metabolism that is needed for short sprints to the aerobic metabolism used to supply energy for long-distance races. Table 1 shows values for the two exponents β_{an} and β_{ae} (for anaerobic and aerobic, respectively) for men and women in athletics at different epochs. This transition is consistently evident since athletics records began.

The slopes of the two scaling laws can be used as a test to compare how evolved and how close to the limit performances are for athletes of different countries of origin, age or sex, for example. The smaller the scaling exponents, the better the performance in longer races, in the sense that the dissipated power will decrease more slowly. In running, men and women show comparable efficiency (their exponents are the same within error): the commonly held belief that women are better than men in long-distance races is not confirmed by our analysis. In swimming, small differences probably result from a difference in buoyancy, which helps women's swimming performance. The slopes are significantly steeper for running than for swimming, perhaps because swimming demands a small aerobic contribution for short races.

Sandra Savaglio*, Vincenzo Carbone†

*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA
e-mail: savaglio@stsci.edu

†Dipartimento di Fisica and Istituto Nazionale di Fisica per la Materia, Unità di Cosenza, Università della Calabria, I-87036 Rende (CS), Italy

1. Katz, J. S. & Katz, L. J. *Sports Sci.* **17**, 467–476 (1999).
2. Prampero, P. E. *Rev. Physiol. Biochem. Pharmacol.* **89**, 143–222 (1981).
3. Moggi, P., Laforluna, C., Russo, G. & Minetti, A. *Eur. J. Appl. Physiol.* **49**, 287–299 (1982).
4. Kennelly, A. E. *Proc. Am. Acad. Arts Sci.* **42**, 273–331 (1906).
5. Henry, F. M. *Res. Q.* **26**, 147–158 (1955).

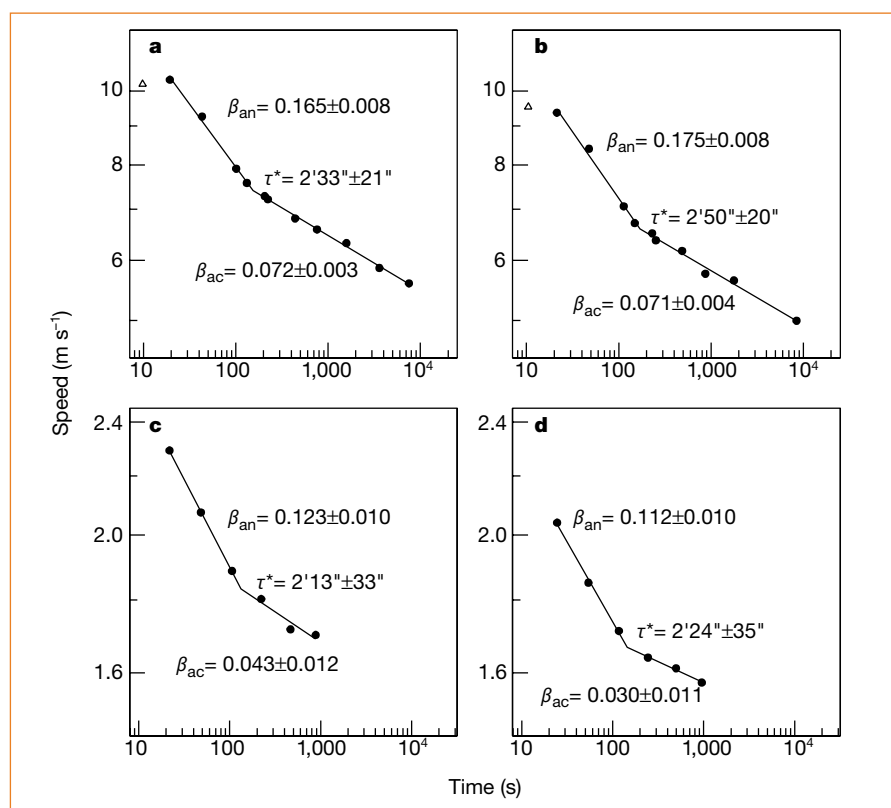


Figure 1 Plots of world-record mean speeds against the record time (at November 1999). **a,b**, Running, and **c,d**, swimming records: for men (**a,c**), we consider 11 races (200 m, 400 m, 800 m, 1,000 m, 1,500 m, the mile, 3,000 m, 5,000 m, 10,000 m, 1 hour, and marathon); the same races are considered for women (**b,d**), apart from the 1 hour race. Lines represent the best fits. The scaling exponents β and characteristic times τ^* of the breakpoints are shown; characteristic times have been determined by using a χ^2 minimization on a broken power law. Triangles in **a,b** represent the 100 m race, which is excluded from the analysis because the mean speed is strongly affected by the standing start of athletes.

Transcription

Gene silencing in worms and fungi

The introduction into cells of foreign nucleic acid molecules can induce sequence-specific gene silencing in some organisms. Here we show that two distantly related organisms, the nematode *Caenorhabditis elegans* and the fungus *Neurospora crassa*, which have quite different mechanisms of gene silencing, both use a similar protein to control the process. This suggests that they may share an ancestral mechanism that evolved to protect the genome against invasion by foreign DNA.

In homology-dependent gene silencing (HDGS), gene expression can be prevented by either DNA or RNA molecules. For example, double-stranded (ds) RNA can inhibit gene expression¹ in *C. elegans*, *Drosophila* and *Trypanosoma*, and transgenic DNA can induce gene silencing in plants² and fungi³.

In a process known as dsRNA interference, dsRNA silences the expression of endogenous *C. elegans* genes after they have been transcribed by inducing sequence-specific degradation of homologous messenger RNA molecules⁴, preventing their translation into protein. Plants and fungi also use post-transcriptional gene silencing (PTGS), but this is induced by transgenic DNA. Again, transcription of the target gene is unaffected but its transcripts do not accumulate as a consequence of rapid degradation.

There are several indications that RNA intermediates are involved even in transgene-induced gene silencing — transgenes produce RNA molecules known as aberrant RNAs, which somehow manage to induce messenger RNA degradation⁵. The two mechanisms of PTGS and dsRNA interference both involve sequence-specific degradation of homologous messenger RNA and an ability of silencing to spread from cell to cell.

Although HDGS phenomena have been considered to be related mechanistically and evolutionarily, no experimental evidence has been forthcoming to support this idea, principally because so far it has not been possible to compare the underlying genetic mechanisms. Genetic characterization of PTGS has been reported for *Arabidopsis thaliana*⁶ and *N. crassa*⁶, and mutants unable to perform dsRNA interference have been isolated from *C. elegans*^{7,8}. These mutants should enable the different components of the cell's gene-silencing machinery to be identified. By comparing the sequences of the respective protein products, we have found that the *qde-2* gene, which controls transgene-induced gene silencing in *N. crassa*⁶, is homologous

QDE-2	T V A N Y Y K Q R Y G I T A N A S L P L V N V G T K E K A I Y V L A E F C T L - V K G R S V K A K L T A N E A D N M I K	433
RDE-1	T L F K I Y E E N K K F I E F P H L P L V K V K S G A K E Y A V P M E H L E V H E K P Q R Y K N R I D L V M Q D K F L K	439
QDE-2	F A C R A P S L N A Q S I Y T K G R Q T L G D K S L T - L G K F K V S I D K E L I T V V G R E L K P P M L T Y S S N K	492
RDE-1	R A T R K P H D Y K E N T L K M L K E L D F S S E E L N F V E R F G L C S K L Q M I E C P G K V L K E P M L V N S V N E	499
QDE-2	T V E - - P O D G G W L M K F V K V A - - - - - R P C R K - - - T E K M T Y L E L K G S - - - -	526
RDE-1	Q I K M T P V I R G F Q E K Q L N V V P E K E L C C A V F V V N E T A G N P C L E E N D V V K F - Y T E L I G G C K F R	558
QDE-2	- - - - K A N E G V P Q A M T A F A E F L N R T G I P I N P R F S P G M S - M S V P G S E - K E F F A K V K E L M S S H	580
RDE-1	G I R I G A N E N R G A Q S I M Y D A T K E Y A F Y K N C T L N T G I R F E I A A T E A K N M F E R L P D - K E Q K	617
QDE-2	Q F V V L L P R K D V A I Y N M V K R A A D I T F G V - - - - - H T V C C A E K F L S T K Q L G Y F A N V G L	633
RDE-1	V L M F I I S K R Q L N A Y G F V K H Y C D H T I G V A N Q H I T S E T V T K A L A S L R H E K G S K R I F Y Q I A L	677
QDE-2	K V N L K - E G G I N H N I - - - - - K T P T P L L A K G K T M V V G Y D V T H P T N L A A G Q S P A	678
RDE-1	K I N A K L G G I N Q E L D W S E I A E I S P E E K E R R K T M P L - - - - - T M Y V G I D V T H P T S Y S G I D Y - -	730
QDE-2	S A P S T I V G L V S T I D Q H L G Q W P A V M W N N P - - - - - H G Q E S M T E Q F T D K F K T R L E L W R S	728
RDE-1	- - - S I A A V A S I N P G G T I Y R N L V T Q E C R P G E R A V A H G R E R - T D I L E A K F V K L R E F A E	786
QDE-2	N P A N N R S L P E N T L I F R D G V S E G F Q M V I K D E L P L V R A A C K L V Y P - - - A G K L P R I T L I V S	784
RDE-1	N - - N D N R A P A H I V V Y R D G V S D S E M L R V S H D E L R S L K S E V I Q F M S E R D G E D P E P K Y T F I V I	844
QDE-2	V K R H Q T R F F P T D P K I H I F K S K - - - - - S P K E G T V V D	814
RDE-1	Q K R H N T R L L R M E K O K P V V N K O L T P A E T D V A V A A V K Q W E E D M K S E T G I V N P S S G T I V D	904
QDE-2	R G V T N V R Y D F F L Q A H A S L Q G T A R S A H Y T V L V D E I F R A D Y G N K A A D T L E Q L T H D M C Y L F G	874
RDE-1	K L I V S K Y K F D F F L A S H H G V L G T S R P G H Y T V M Y D D - - - - - K G M S Q D E V Y K M T Y G L A F L S A	958
QDE-2	R A T K A V S I C P P A Y A D L V C D R A R I H Q K E L F D A L D E N	910
RDE-1	R C R K P T S L P V P V H Y A H L S C E K A - - - - K E L Y R T Y K E H	990

Figure 1 Sequence alignment of the QDE-2 protein (Genbank accession number AF217760) with RDE-1 protein (AF180730) of *C. elegans*. The conserved carboxy-terminus region (expected value 10^{-23} , 38% similarity) is shown. Identical residues are in black; conservative substitutions are in grey.

to the *rde-1* gene⁷, which is essential for dsRNA interference in *C. elegans* (Fig. 1).

This finding is, to our knowledge, the first experimental evidence indicating that dsRNA interference and PTGS induced by transgenic DNA share a common genetic mechanism. It supports the idea that HDGS phenomena evolved from an ancestral mechanism aimed to protect the genome against transposons and viruses. Our results also suggest that dsRNA molecules might participate in PTGS in fungi.

dsRNA could be produced directly from integrated transgenes as a result of the presence of inverted repeats, or as an outcome of transcription from convergent inverted promoters. Alternatively, transgenic single-stranded aberrant RNA may be used as a template by QDE1, a putative RNA-dependent RNA polymerase⁹, to produce dsRNAs. Further analysis should define the mechanistic function of *rde-1* and *qde-2* in gene silencing and confirm whether the homology between these two genes effectively corresponds to a cognate step in the two gene-silencing mechanisms.

As well as including *rde-1*, the *qde-2* gene family has members in both plants

and animals: in *A. thaliana* and *Drosophila*, *qde-2* homologues have been implicated in the regulation of development¹⁰, and the rabbit *qde-2* homologue, eIF2C, forms part of a protein complex that stimulates the start of translation¹¹. As all the *qde-2* homologues have not yet been biochemically characterized, it is possible that these evolutionarily related genes could be required for different biological processes.

Caterina Catalanotto, Gianluca Azzalin, Giuseppe Macino, Carlo Cogoni

Dipartimento di Biotecnologie Cellulari ed Ematologia, Sezione di Genetica Molecolare, Università di Roma La Sapienza, Viale Regina Elena, 324, 00161 Roma, Italy
e-mail: carlo@bce.med.uniroma1.it

1. Fire, A. *Trends Genet.* **15**, 358–363 (1999).
2. Wassenaar, M. & Pelissier, T. *Plant Mol. Biol.* **37**, 349–362 (1998).
3. Cogoni, C. & Macino, G. *Trends Plant Sci.* **2**, 438–443 (1997).
4. Montgomery, M. K., Xu, S. & Fire, A. *Proc. Natl Acad. Sci. USA* **95**, 15502–15507 (1998).
5. Elmayan, T. et al. *Plant Cell* **10**, 1747–1758 (1998).
6. Cogoni, C. & Macino, G. *Proc. Natl Acad. Sci. USA* **94**, 10233–10238 (1997).
7. Tabara, H. et al. *Cell* **99**, 123–132 (1999).
8. Ketting, R. F., Haverkamp, T. H., van Luenen, H. G. & Plasterk, R. H. *Cell* **99**, 133–141 (1999).
9. Cogoni, C. & Macino, G. *Nature* **399**, 166–169 (1999).
10. Benfey, P. N. *Curr. Biol.* **9**, 171–172 (1999).
11. Zou, C., Zhang, Z., Wu, S. & Osterman, J. C. *Gene* **211**, 187–194 (1998).

Brief communications are intended to provide a forum for short, topical reports of general scientific interest and for technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on Nature's website (www.nature.com) or on request from nature@nature.com

Quantum information and computation

Charles H. Bennett & David P. DiVincenzo

IBM Research Division, T. J. Watson Research Center, Yorktown Heights, New York 10598, USA

In information processing, as in physics, our classical world view provides an incomplete approximation to an underlying quantum reality. Quantum effects like interference and entanglement play no direct role in conventional information processing, but they can—in principle now, but probably eventually in practice—be harnessed to break codes, create unbreakable codes, and speed up otherwise intractable computations.

Information and computation theory have undergone a spurt of new growth, and a renewal of their historic connection to basic physics, as they have expanded to treat the intact transmission and processing of quantum states, and the interaction of such 'quantum information' with traditional forms of information. We may wonder why this did not happen earlier, as quantum principles have long been accepted as fundamental to all of physics. Perhaps the founders of information and computation theory, such as Shannon, Turing and von Neumann, were too accustomed to thinking of information processing in macroscopic terms, not yet having before them the powerful examples of the genetic code and ever-shrinking microelectronics. Be that as it may, information until recently has largely been thought of in classical terms, with quantum mechanics playing a supporting role in the design of the equipment to process it, and setting limits on the rate at which it could be sent through certain channels. Now we know that a fully quantum theory of information and information processing offers, among other benefits, a brand of cryptography whose security rests on fundamental physics, and a reasonable hope of constructing quantum computers that could dramatically speed up the solution of certain mathematical problems. These benefits depend on distinctively quantum properties such as uncertainty, interference and entanglement.

At a more fundamental level, it has become clear that an information theory based on quantum principles extends and completes classical information theory, just as complex numbers extend and complete the reals. Besides quantum generalizations of classical notions such as sources, channels and codes, the new theory includes two complementary, quantifiable kinds of information—

classical information and quantum entanglement. Classical information can be copied at will, but can only be transmitted forward in time, to a receiver in the sender's forward light cone. Entanglement in contrast, cannot be copied, but can connect any two points in space-time. Conventional data processing operations destroy entanglement, but quantum operations can create it and use it for various purposes, such as speeding up certain classical computations and assisting in the transmission of classical information or intact quantum states. Part of the new quantum information theory is the qualitative and quantitative study of entanglement, and its interactions with classical information.

Any means, such as an optical fibre, of delivering quantum systems more or less intact from one place to another, may be viewed as a quantum channel. Unlike classical channels, which are well characterized by a single capacity, quantum channels have several distinct capacities, depending on what one is trying to use them for, and what auxiliary resources are brought into play.

New effects involving quantum information continue to be discovered, not only in the traditional areas of computation, channel capacity, and cryptography, but in areas such as communication complexity and game theory.

Theory of quantum data and data processing

Quantum data. How, then, does quantum information, and the operations that can be performed on it, differ from conventional digital data and data-processing operations? A classical bit (such as a memory element or a wire carrying a binary signal) is generally a macroscopic system, and is described by one or more continuous parameters such as voltages. Within this parameter space

Table 1 Comparison of classical and quantum information processing

Property	Classical	Quantum
State representation	String of bits $x \in \{0, 1\}^n$	String of qubits $\psi = \sum_i c_i x_i\rangle$
Computation primitives	One- and two-bit boolean operations	One- and two-qubit unitary transformations
Fault-tolerant computation	By classical fault-tolerant gate arrays	By quantum fault-tolerant gate arrays
Quantum computational speed-ups		Factoring: exponential speed-up; search: quadratic speed-up; iteration, parity: no speed-up; simulation of quantum systems: up to exponential speed-up
Communication primitives	Transmitting a classical bit	Transmitting a classical bit; transmitting a qubit; sharing an EPR pair
Noiseless coding techniques	Classical data compression	Quantum data compression; entanglement concentration
Error-correction techniques	Error-correcting codes	Quantum error-correcting codes; entanglement distillation
Noisy-channel capacities	Classical capacity C_1 equals maximum mutual information through a single channel use	Classical capacity $C \geq C_1$; unassisted quantum capacity $Q \leq C$; classically assisted quantum capacity $Q_2 \geq Q$; entanglement-assisted classical capacity $C_E \geq C$
Entanglement assisted communication		Superdense coding, quantum teleportation
Communication complexity	Bit communication cost of distributed computation	Qubit cost, or entanglement-assisted bit cost, can be less
Secret cryptographic key agreement	Known protocols insecure against quantum computer	Secure against general quantum attack and unlimited computing
Two-party bit commitment	Known protocols insecure against quantum computer	Insecure against attack by a quantum computer

two well-separated regions are chosen by the designer to represent 0 and 1, and signals are periodically restored toward these standard regions to prevent them from drifting away due to environmental influences, crosstalk, and finite manufacturing tolerances. An n -bit memory can exist in any of 2^n logical states, labelled 000...0 to 111...1. Besides storing binary data, classical computers manipulate it; a sequence of boolean operations (for example, NOT and AND) acting on the bits one or two at a time is sufficient to realize any deterministic transformation.

A quantum bit or 'qubit' in contrast, is typically a microscopic system, such as an atom or nuclear spin or photon. The boolean states 0 and 1 are represented by a fixed pair of reliably distinguishable states of the qubit (for example, horizontal and vertical photon polarizations: $|0\rangle = \leftrightarrow$, $|1\rangle = \updownarrow$). A qubit can also exist in a continuum of intermediate states or 'superpositions', represented mathematically as complex linear combinations of the basis states $|0\rangle$ and $|1\rangle$. For photons, these states correspond to other polarizations, for example $\nearrow = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, $\nwarrow = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $\curvearrowright = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle)$ (right circular polarization). Unlike the intermediate states of a classical bit (such as voltages between the standard 0 and 1 values), these intermediate states cannot be reliably distinguished, even in principle, from the basis states. With regard to any measurement, the superposition $\alpha|0\rangle + \beta|1\rangle$ behaves like $|0\rangle$ with probability $|\alpha|^2$ and like $|1\rangle$ with probability $|\beta|^2$. More generally two quantum states are reliably distinguishable if and only if their vector representations are orthogonal; thus $\leftrightarrow$ and $\updownarrow$ are reliably distinguishable by one type of measurement, and $\nearrow$ and $\nwarrow$ by another, but no measurement can reliably distinguish $\leftrightarrow$ from $\nearrow$.

A pair of qubits (for example, two photons in different locations) is capable of existing in four boolean states, $|00\rangle$, $|01\rangle$, $|10\rangle$ and $|11\rangle$, as well as all possible superpositions of them. These include states such as

$$\frac{1}{\sqrt{2}}(|00\rangle + |01\rangle) = |0\rangle \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) = \leftrightarrow \nearrow \quad (1)$$

which is describable as a tensor product of states of the individual photons, as well as states such as $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ which admit no such description. Such 'entangled' states correspond to a situation in which neither photon by itself has a definite state, even though the pair together does.

More generally, where a string of n classical bits could exist in any of 2^n boolean states $x = 000...0$ through $111...1$, a string of n qubits can exist in any state of the form

$$\Psi = \sum_{x=00...0}^{11...1} c_x |x\rangle \quad (2)$$

where the c_x are complex numbers such that $\sum_x |c_x|^2 = 1$. In other words, a quantum state of n qubits is represented by a complex vector Ψ of unit length in a space ('Hilbert space') of 2^n dimensions, one for each possible classical state. The exponentially large dimensionality of this space distinguishes quantum computers from classical analogue computers, whose state is described by a number of parameters that grows only linearly with the size of the system. This is because classical systems, whether digital or analogue, can be completely described by separately describing the state of each part. The vast majority of quantum states, by contrast, are entangled, and admit no such description. The ability to preserve and manipulate entangled states is the distinguishing feature of quantum computers, responsible both for their power and for the difficulty of building them.

An isolated quantum system evolves in such a way as to preserve superpositions and distinguishability; such evolution, called 'unitary', is the Hilbert-space analogue of rigid rotation in real space, and is another important difference between quantum and analogue systems. Unitary evolution and superposition are the central principles of quantum mechanics.

Logical operations. Just as any classical computation can be expressed as a sequence of one- and two-bit operations (for example, NOT and AND gates), any quantum computation can be expressed as a sequence of one- and two-qubit quantum gates, that is, unitary operations acting on one or two qubits at a time¹ (compare with Fig. 1). The most general one-qubit gate is described by a 2×2 unitary matrix $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ mapping $|0\rangle$ to $\alpha|0\rangle + \beta|1\rangle$ and $|1\rangle$ to $\gamma|0\rangle + \delta|1\rangle$. One-qubit gates can easily be implemented physically, for example, by quarter- and half-wave plates acting on polarized photons, or by radio-frequency tipping pulses acting on nuclear spins in a magnetic field.

The standard two-qubit gate is the controlled-NOT or XOR gate, which flips its second (or 'target') input if its first ('control') input is $|1\rangle$ and does nothing if the first input is $|0\rangle$. In other words it interchanges $|10\rangle$ and $|11\rangle$ while leaving $|00\rangle$ and $|01\rangle$ unchanged. Unlike one-qubit gates, two-qubit gates are difficult to realize in the laboratory, because they require two separated quantum information carriers to be brought into strong and controlled interaction.

The XOR gate is a prototype interaction between two quantum systems, and illustrates several key features of quantum information, in particular the impossibility of cloning an unknown quantum state, and the way interaction produces entanglement. If the XOR is applied to boolean data in which the second qubit is 0 and the first is 0 or 1, the effect is to leave the first qubit unchanged while the second becomes a copy of it: $U_{\text{XOR}}|x, 0\rangle = |x, x\rangle$ for $x = 0$ or 1 . One might suppose that the XOR operation could also be used to copy superpositions, such as $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, so that $U_{\text{XOR}}|\psi, 0\rangle$ would yield $|\psi, \psi\rangle$, but this is not so. The unitarity of quantum evolution requires that a superposition of input states evolve to a corresponding superposition of outputs. Thus the result of applying U_{XOR} to $|\psi, 0\rangle$ must be $\alpha|0, 0\rangle + \beta|1, 1\rangle$, an entangled state in which neither output qubit alone has a definite state. If one of the entangled output qubits is lost (for example, discarded, or allowed to escape into the environment), the other thenceforth behaves as if it had acquired a random classical value 0 (with probability $|\alpha|^2$) or 1 (with probability $|\beta|^2$). Unless the lost output is brought back into play, all record of the original superposition $|\psi\rangle$ will have been lost. This behaviour is characteristic not only of the XOR gate but of unitary interactions generally: their typical effect is to map most unentangled initial states of the interacting systems into entangled final states, which from the viewpoint of either system alone causes an unpredictable disturbance.

Environmental interactions. Since quantum physics underlies classical, there should be a way to represent classical data and operations within the quantum formalism. If a classical bit is a qubit having the value $|0\rangle$ or $|1\rangle$, a classical wire should be a wire that conducts $|0\rangle$ and $|1\rangle$ reliably, but not superpositions. This can be implemented using the XOR gate as described above, with an initial $|0\rangle$ in the target position which is later discarded. In other words, from the viewpoint of quantum information, classical communication is an irreversible process in which the signal interacts *en route* with an environment in such a way that boolean signals pass through undisturbed, but other states suffer entanglement with the environment. If the environment is lost or discarded, the surviving signal behaves as if it had been irreversibly forced to choose one of the boolean states. Not only a classical wire, but any classical data processing, can be realized similarly by quantum processing supplemented by interaction with a quantum environment that is later discarded.

Paradoxically, entangling interactions with the environment are thought to be the main reason why the macroscopic world seems to behave classically and not quantum-mechanically². Macroscopically different states, for example, the different charge states representing 0 and 1 in a VLSI (very large scale integration) memory cell, interact so strongly with their environment that information rapidly leaks out as to which state the cell is in. Therefore, even if it were possible to prepare the cell in a superposition of 0 and 1, the superposition

would rapidly evolve into a complex entangled state involving the environment, which from the viewpoint of the memory cell would appear as a statistical mixture, rather than a superposition, of the two classical values. The spontaneous decay of superpositions into mixtures is known as decoherence.

Entanglement with the environment is thus a major obstacle to quantum computation. To avoid having a quantum computation decohere into a probabilistic classical computation (which could just as well be done on a classical computer) it is necessary, while creating and maintaining entanglement among the computational degrees of freedom, to avoid entanglement between them and the environment. Until recently it appeared that the feasible number of steps in a coherent quantum computation would necessarily be less than the ratio τ_d/τ_s of decoherence time to switching time characteristic of the elementary quantum systems used in the hardware. Even if all other problems in the design of a practical quantum computer could be overcome, currently attainable values of τ_d/τ_s are not high enough to make quantum computers competitive with classical ones; also, the search for systems with ever-higher τ_d/τ_s might ultimately be blocked by fundamental properties of available atoms and nuclei. Apart from decoherence, it also appeared that individual gate operations would have to be made more and more precise the longer the computation.

This pessimism has largely been dispelled by the discovery of quantum fault-tolerant computation³ (QFTC), the quantum analogue of von Neumann's discovery that unreliable classical gates can

be used to perform arbitrarily long classical computations reliably, provided the error probability per gate is less than some constant threshold. Because of QFTC, it appears that experimentalists need 'only' build quantum hardware with a per-gate decoherence that is below some finite threshold (variously estimated at 10^{-6} to 10^{-2} , with a similar precision for individual gate rotations) in order for quantum computers to do arbitrarily complex computations.

With this background we survey some of the main parallels and differences between quantum and classical information processing. **Quantum speed-up of classical computation.** This is potentially the most important application of quantum data processing. By using quantum gates and wires, with entangled states flowing through them in the intermediate stages of a computation, certain computations mapping classical inputs x to classical outputs $f(x)$ can be done in far fewer steps than by any known sequence of classical gate operations. Most famously, a quantum computer can factor large integers in time that is polynomial in the logarithm of the best classical time^{4,5}, thereby threatening the security of cryptosystems based on the presumed difficulty of factoring. This exponential speed-up depends on the quantum computer's ability to vastly parallelize the performance of a fast Fourier transform, using destructive interference among a number of parallel computation paths that increases exponentially with the number of physical qubits involved in the computation. Another class of problems for which quantum computers seem to provide exponential speed-up is the simulation of many-particle quantum systems^{6,7}. In contrast to these rather specialized problems, a much broader class of problems can be speeded up quadratically, that is, solved in a time proportional to the square root of the time that a classical computer would require. These include search and optimization problems (for example, given an algorithm for computing a function F , find an input s where $F(s) = 0$, or an input s where $F(s)$ is a minimum)^{8,9}. For some other problems there is no quantum speed-up. These include iterated function evaluation^{10,11} (for example, given an algorithm for computing F , compute the n th iterate $F^{(n)} = F(F(\dots F(\dots)))$ for large n) and computing the parity of a random set^{12,13}.

Quantum information theory. This generalizes the classical notions of source and channel, and the related techniques of source and channel coding, as well as introducing a new resource, entanglement, which interacts with classical and quantum information in a variety of ways that have no classical parallel.

As mentioned earlier, quantum channels have several distinct capacities, depending on what one is trying to use them for, and what auxiliary resources are brought into play. These include the following:

Classical capacity, C , equal to the maximum rate at which classical bits can be transmitted reliably through the channel;

Quantum capacity, Q , the maximum rate at which intact qubits can be transmitted reliably through the channel;

Classically-assisted quantum capacity, Q_2 , defined as the maximum rate at which qubits can be transmitted reliably through the channel, with the help of unlimited two-way classical communication between sender and receiver; and

Entanglement-assisted classical capacity, C_E , defined as the maximum rate for sending classical bits through the channel, with the help of unlimited prior entanglement between sender and receiver. These capacities obey the relation $Q \leq Q_2 \leq C \leq C_E$ for all known channels, but otherwise appear to vary rather independently, and are not easy to calculate from the quantum channel parameters, again, unlike the single capacity of classical channels.

Quantum data compression and error correction. The two central techniques of classical information theory, source and channel coding, have direct quantum analogues; a quantum source is an entity that emits quantum states ψ_i with probabilities p_i , and a channel is an entity, such as an optical fibre, that transmits quantum states more or less reliably from a sender to a receiver.

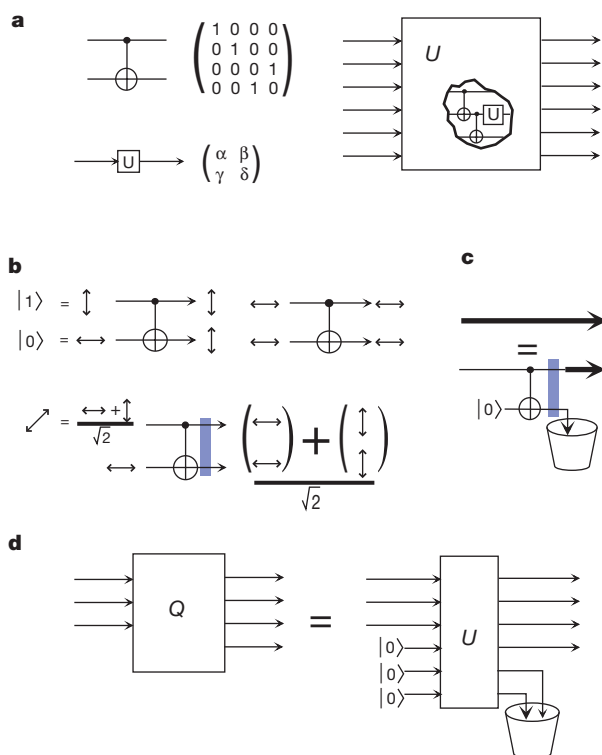


Figure 1 Quantum logical operations. **a**, Any unitary operation U on quantum data can be synthesized from the two-qubit XOR or controlled-NOT gate, and one-qubit unitary operations U . **b**, The XOR acts as a classical cloner on boolean valued inputs, but if one attempts to clone intermediate values, the cloning fails and an entangled state (blue) results instead. **c**, A classical wire (thick line) conducts 0 and 1 faithfully but not superpositions or entangled states. It may be defined as a quantum wire that interacts (via an XOR) with an ancillary 0 qubit which is then discarded. **d**, The most general treatment, or superoperator, Q that can be applied to quantum data is a unitary interaction with one or more 0 qubits, followed by discarding some of the qubits. Superoperators are typically irreversible.

The von Neumann entropy of a quantum source, $S = -\text{Tr}(\rho \log_2 \rho)$, where $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, determines the minimum asymptotic number of qubits into which its signals can be compressed by a quantum encoder and still be faithfully recovered by a quantum decoder. This is the analogue of classical data compression or source coding, by which redundant classical data is compressed and faithfully regenerated, but quantum data compression¹⁴ differs in that it can be applied to non-orthogonal states (for example, equally probable horizontal and diagonal photons, as shown in Fig. 2a) which would be spoiled if one tried to compress them classically. Also, because the states are non-orthogonal, the encoder cannot retain a copy of them, or indeed any memory of them, if they are to be faithfully reconstructed at the receiving end. A quantum encoder is like a discreet telegrapher, who transmits messages without remembering them.

Source coding removes redundancy, allowing data to be sent more efficiently through a noiseless channel. Error-correction or channel coding, in contrast, introduces redundancy to enable data to withstand transmission through a noisy channel. The simplest classical error-correcting code is the triple repetition code $0 \rightarrow 000$, $1 \rightarrow 111$, which permits the encoded bit to be faithfully recovered after up to one transmission error in the three-bit codeword. Analogous error-correcting codes exist for quantum data, but they require more redundancy because they need to protect not only boolean states, but also arbitrary superpositions of them^{15–20}. Thus the simplest single-error-correcting quantum code (Fig. 2b) encodes an arbitrary input qubit $|\psi\rangle$ into an entangled state of five qubits, in such a way that if any one is corrupted *en route*, the decoder can funnel the effects of the error into the four ancillary qubits, while restoring the first qubit to its original state. Analogously to classical capacity, the quantum capacity Q of a noisy channel can be defined as the limiting ratio of faithfully transmitted qubits per noisy-channel use that is achievable by quantum error-correcting codes. This quantum capacity is usually less, and can never be greater, than the same channel's capacity C for transmitting classical bits. The inequality $Q \leq C$ holds for all channels because if a channel can faithfully transmit a general qubit, then it can certainly transmit the particular qubits $|0\rangle$ and $|1\rangle$.

The discovery of quantum error-correcting codes in 1995 came as a great surprise, probably because people were used to thinking of classical error correction in language unsuited to quantum generalization. For example, triple repetition, if it is taken to mean making three copies of the input qubit ($\psi \rightarrow \psi \otimes \psi \otimes \psi$), flies in the face of the well-known impossibility of exactly copying ('cloning') an unknown quantum state. In retrospect, the natural quantum generalization of triple repetition can be seen instead to be the mapping $\alpha|0\rangle + \beta|1\rangle \rightarrow \alpha|000\rangle + \beta|111\rangle$, which does not violate any quantum principle and indeed suffices to correct single qubit errors in any boolean input. As noted above, two more bits of redundancy are needed to extend the protection to non-boolean inputs. Although analogous in structure to classical discrete error-correcting codes, quantum error-correcting codes have the remarkable ability to protect a continuum of inputs from a continuum of errors. For example, in Fig. 2b, the input qubit might be a photon of any polarization state, and the error (red) might be a rotation of one of the five channel qubits' polarizations by an arbitrary amount; nevertheless, the error would be corrected. This is a beneficial side effect of the linearity of quantum mechanics: if a quantum error-correcting code protects a sufficiently rich discrete set of inputs from a sufficiently rich discrete set of error processes, then it will also protect any superposition of those inputs from any superposition of those errors. Besides the simple capacities C and Q , quantum channels have assisted capacities Q_2 and C_E mentioned above, which will be discussed later.

The oldest branch of quantum information theory^{21–23} concerns the use of quantum channels to transmit classical information. Even the seemingly pedestrian classical capacity C is not easy to calculate

for quantum channels, because it may depend on using a quantum encoder to prepare inputs entangled over multiple uses of the channel, and/or a quantum decoder to perform coherent measurements on multiple channel outputs. Unlike any classical channel, some quantum channels are superadditive, in the sense that more classical information can be sent through n parallel uses of the channel than n times the amount that can be sent through one use of the channel^{24–28}.

Entanglement-assisted communication. Two forms of quantum information transmission that have no classical counterpart, but are closely related to each other, are quantum teleportation²⁹ (Fig. 3a) and quantum superdense coding³⁰ (Fig. 3b). These involve an initial stage in which a pair of particles in a maximally-entangled state such as $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ (often called an Einstein–Podolsky–Rosen or EPR pair) is shared between two parties, followed by a second stage in which this shared entanglement is used to achieve, respectively, transmission of a qubit via two classical bits, or transmission of two classical bits via one qubit. Quantum teleportation illustrates the fact that transmission of intact quantum states requires two qualitatively different resources: a quantum resource that cannot be cloned, and a directed resource that cannot travel faster than light. In direct transmission of a qubit, these two functions are performed by the same particle. In teleportation the former function is provided by the shared EPR pair, the latter by the two classical bits. This situation may be summarized by saying that classical information theory involves one species of information, and one kind of noiseless communication primitive (transmission of a bit), whereas quantum information theory involves two species (classical information and entanglement), and three primitives (transmitting a bit, transmitting a qubit, and sharing an EPR pair) which are related through superdense coding and teleportation.

Superdense coding (Fig. 3b) is an example of entanglement-assisted classical communication, and shows that $C_E = 2$ for the noiseless qubit channel, while $C = Q = 1$. Surprisingly, the ratio

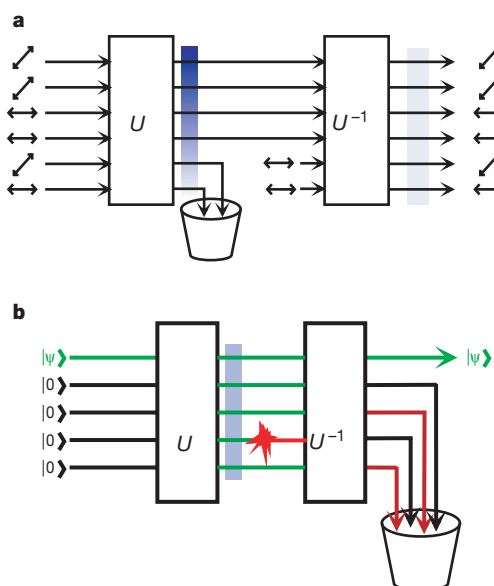


Figure 2 Quantum data compression and error correction. **a**, In quantum data compression, inputs from a redundant source (here, an unknown sequence of horizontal and diagonal photons) are unitarily transformed into an entangled state (blue) in which almost all the information has been concentrated into some of the photons, allowing the others to be discarded. At the receiving end of the channel the discarded photons are replaced by standard (horizontal) photons and the unitary transformation is undone, resulting in a close approximation to the original state. **b**, A quantum error-correcting code with unitary encoder and decoder. An arbitrary input qubit $|\psi\rangle$ is entangled with four standard $|0\rangle$ qubits in such a way that if any one of the five qubits is spoiled, the decoder can still restore the original state exactly.

C_E/C typically increases with increasing noise, and indeed can attain arbitrarily large values for channels so noisy that their quantum capacities Q and Q_2 both vanish³¹. Thus, unlike most quantum effects, entanglement enhancement of a quantum channel's classical capacity does not disappear in the limit of large noise. In this respect it resembles the ability of bulk nuclear magnetic resonance systems to carry out nontrivial quantum computations while remaining close to thermal equilibrium.

Superdense coding and teleportation have received much laboratory attention recently. The first work was by the Innsbruck group³², which implemented a version of superdense coding in which three distinguishable states (rather than the theoretical maximum of four) are created by manipulating one member of an EPR pair of polarization-entangled photons. Teleportation using these photon states was more recently achieved by the same group³³; by using these techniques several other protocols involving entanglement, for example, the creation of three-particle entanglement, has become possible. An experimentally different approach in which another attribute of one of the EPR photons (such as its position) is teleported has been implemented in Rome³⁴. This experiment is easier in that it involves two rather than three photons. A very recent experiment³⁵ has followed more directly a version of a teleportation scheme due to Vaidman³⁶, in which continuous quantum degrees of freedom are teleported. This work demonstrates that an arbitrary quantum state of one optical mode can be teleported with good fidelity; taking into account limitations on the intensity of the mode, one finds that roughly a one-million-state quantum system is involved, in contrast to the two-state systems used in the earlier work. Finally, the operations necessary to teleport a nuclear spin state have been performed using nuclear magnetic resonance³⁷, although the range of teleportation is only the distance across a single molecule.

Although entanglement by itself cannot be used to transmit a classical message, it can reduce the amount of classical communication required to perform a distributed computation^{13,38,39}. Classically, 'communication complexity' refers to the amount of communication needed to evaluate a function of several inputs in remote locations. For example, if Alice and Bob each have appointment calendars with n time slots, $O(n)$ bits of communication are required to determine if there is a time when they are both free. If they are allowed to share prior entanglement, or if they are allowed to communicate using qubits rather than bits, the communication complexity of this problem is reduced from $O(n)$ to $O(\sqrt{n} \log n)$.

Quantifying and distilling entanglement. Because of its usefulness in protocols such as teleportation, it is important to have quanti-

tative measures of entanglement, and to know whether all entangled states (those not expressible as products of states of their parts, or probabilistic mixtures of such products) can be converted into EPR pairs, and if so, how efficiently. In the case of bipartite pure states, entanglement is naturally measured by the state's entropy of entanglement, the von Neumann entropy of either subsystem considered alone. For such states^{40–42} the entropy of entanglement $E(\Psi)$ is equal both to the state's entanglement of formation—the number of EPR pairs asymptotically required to prepare one instance of the state by classical communication and local operations—and its distillable entanglement—the number of pure EPR pairs that can be asymptotically prepared from one instance of the state by classical communication and local operations.

For mixed states, and states of three or more parties, the situation is more complicated, and there are several non-equivalent kinds of entanglement. Multipartite states, pure and mixed, have been studied^{43–45}. Mixed states generally have a distillable entanglement that is less than their entanglement of formation, reflecting the irreversibility of the mixing process. An extreme form of this phenomenon is the existence of so-called "bound" entangled states⁴⁶—mixed states which are entangled, but from which no pure entanglement can be distilled.

Entanglement distillation is important not only for quantifying entanglement but as a distinctively quantum kind of error correction, complementary to the use of quantum error-correcting codes^{20,47,48}. Suppose Alice and Bob can communicate classically, and in addition have access to a noisy quantum channel. Now Alice wants to send an unknown qubit reliably to Bob. If the quantum channel is not too noisy, she can encode the input qubit into several qubits using a quantum error-correcting code as in Fig. 2b, send these through the noisy channel, and have Bob decode them. However for very noisy channels, such as a 50% depolarizing channel, this will not work, because such channels have zero quantum capacity $Q = 0$. In this case the best known strategy is for Alice not to send the input qubit through the channel at all, but instead prepare a number of pure EPR pairs, and share them through the noisy channel with Bob (resulting in noisy EPR pairs). Then, using their ability to communicate classically, Alice and Bob distill a smaller number of good EPR pairs from the larger supply of noisy ones. Finally, Alice uses one of the good EPR pairs, and additional classical communication, to teleport the input qubit safely to Bob. The ability of entanglement distillation to salvage such noisy EPR pairs gives rise to a fact noted earlier, that for many channels the classically-assisted quantum capacity Q_2 exceeds the direct quantum capacity Q . (However, this advantage depends on two-way classical communication between Alice and Bob—if they are limited to one-way communication, distillation is no more efficient than quantum error-correcting codes.) As a function of increasing noise, a typical quantum channel passes through two thresholds; a noise level beyond which Q vanishes but Q_2 and C remain positive; and a threshold beyond which Q_2 vanishes but C remains positive.

Quantum fault-tolerant computation. QFTC is both an expansion of research in the theory of quantum information processing and a practical necessity for implementing non-trivial quantum computation in the laboratory. Modern QFTC has been well reviewed (see ref. 3 and references therein); some of the basic ideas are sketched in Fig. 4. To avoid irrecoverable damage from a single error, an appropriate quantum error-correcting code is used to spread the logical state $|\psi_L\rangle$ being stored or processed over several physical qubits, carried by a bundle of parallel wires. Periodically the bundle passes through a restorative gate array R , where it interacts with clean ancilla qubits from the environment, in order to correct the errors by funnelling them into the ancillas, which are then discarded. Additional errors may occur during the restoration process itself, but if these are not too numerous they will be corrected by a subsequent restoration step (Fig. 4a). Such a regimen of active

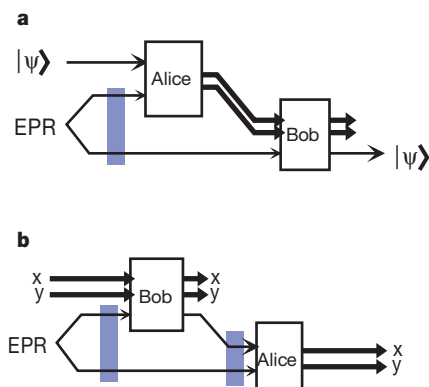


Figure 3 Quantum information transmission between a sender (Alice) and a receiver (Bob). **a**, In quantum teleportation, prior sharing of an EPR pair, and transmission of a two-bit classical message from Alice to Bob suffice to transmit an unknown quantum state even when no direct quantum channel from Alice to Bob is available. **b**, In quantum dense coding, prior sharing of an EPR pair, and transmission of a single qubit from Bob to Alice, suffice to transmit an arbitrary two-bit classical message (x, y).

restoration can be used to implement a fault-tolerant quantum memory, able to hold quantum states reliably for much longer than the natural decoherence times of the hardware of which the array is built. Apart from the fact that it operates on quantum rather than classical data, this is entirely analogous to devices such as the dynamic random access memory (DRAM) used in today's computers, in which periodic signal restoration serves to delay the decay of the stored data almost indefinitely. To perform quantum computation fault-tolerantly, it is also necessary, besides storing the data, to perform gate operations on it without decoding it from its protected form. For some gates, such as the XOR, this can be done in a straightforward fashion, by applying the gate operation successively to wires in a bundle (Fig. 4b). Other gate operations, including some necessary single-qubit rotations, must be implemented in a more complex fashion, involving preparation and testing of special entangled states of a set of ancilla qubits, which are then brought into interaction with the encoded data to perform the desired logical transformation³.

The promise of quantum computation lies in the fact that, to perform a t -step computation fault-tolerantly, the number of gates and wires need only be multiplied by a factor polynomial in $\log t$. Therefore, for computations with a significant quantum speed-up, a quantum computer would still vastly outperform any classical computer on sufficiently large inputs.

Quantum cryptography. This is the art of applying the unique properties of quantum systems to cryptographic goals, that is, the protection of classical information from tampering or unauthorized disclosure in a multi-party setting where not all the parties trust one another. This adversarial element distinguishes it from the kinds of quantum information processing considered earlier.

One important quantum cryptographic task, quantum key distribution has as its goal the sharing of a secret random bit string K , called a cryptographic key, between the two protagonists Alice and Bob, who have at their disposal an insecure quantum channel and a public classical channel. (Purely classical protocols for key agreement exist and are in widespread use, but these result in a key that is not informationally secure—an adversary with sufficient computing power could infer it from the public messages exchanged between Alice and Bob. In particular, the most widely used classical key agreement protocols could be easily broken by a quantum computer, if one were available.) In quantum key distribution, an eavesdropper ('Eve') is allowed to interact with the quantum information carriers (for example, photons) *en route* from Alice to Bob—at the risk of disturbing them—and can also passively listen to all classical communication between Alice and Bob, but she cannot alter or suppress the classical messages. Sometimes (for example, if Eve jams or interacts strongly with the quantum signals) Alice and Bob will detect the excessive eavesdropping and abort the protocol; but, for every eavesdropping strategy, Eve's probability of remaining undetected and obtaining significant information on the key should be negligible.

The practical implementation of quantum key distribution is much farther advanced than other kinds of quantum information processing, owing to the fact that the standard quantum key distribution protocols require no two-qubit interactions, only preparation and measurement of simple quantum states, along with classical communication and computations. Optical prototypes working over tens of kilometres of fibre, or even through a kilometre of open air at night, have been built and tested. In principle, however, a quantum key distribution protocol could involve quantum computations by Alice and Bob; and to be sure of its security, one ought to allow Eve the full power of a quantum computer, even though Alice and Bob do not need one for the standard protocols.

Various proofs of security of quantum key distribution protocols, especially the four-state "BB84" protocol of ref. 49, have been offered. A complete security proof should encompass all attacks

allowed by the laws of quantum mechanics, and should also be able to cope with noise under the realistic assumption that it arises not only from eavesdropping but also from noisy channels and detectors. Finally, it should provide a way of calculating a safe rate of key generation as a function of the noise level observed by Alice and Bob. Recent proofs^{50,51} building on a long history of previous security proofs against more limited attacks^{48,49,52,53}, have largely met these criteria, the main remaining problems being to simplify the proofs, improve the error thresholds, and to extend them to cover realistic sources, which do not emit exact single-photon states or exact EPR pairs, and in extreme cases may even have been sabotaged by Eve.

Given the success of quantum key distribution, there was high hope that quantum techniques could help with another task, two-party oblivious function evaluation, a better name for which might be "discreet decision-making". This is the task, which arises frequently in commerce and diplomacy, of enabling two mutually distrustful parties to cooperate in evaluating a publicly agreed function of private data held separately by each party, without compromising the private data any more than it would have been compromised had they assigned the job of evaluating the function to a trusted intermediary. Initially Alice knows data x and Bob knows data y ; when the protocol is finished, Alice and Bob should each also know $f(x,y)$, but neither party should know any more about the other party's private input than can logically be inferred from a knowledge of their own data and the common function value $f(x,y)$. Classical protocols for oblivious function evaluation exist, but, like classical key agreement protocols, they are not informationally secure and could be broken by a quantum computer. Hopes for finding a quantum basis for absolutely secure oblivious function evaluation were dashed by the discovery that a fundamental building block of all known oblivious function evaluation protocols, called bit commitment, is insecure in principle against quantum attacks^{54,55}. Bit commitment is the idealization of a protocol in which Alice sends Bob a locked box containing a bit 0 or 1 of her choosing, written on a piece of paper, then later, at a time of her choosing, sends him the key so that he can open the box and then read the bit. Quantum bit commitment is insecure because of a fundamental property of entangled states, namely that if two pure states of the Alice–Bob system are indistinguishable to Bob, they must be interconvertible by a local action of Alice; thus there is in principle no way of implementing a locked box containing a bit value that is both unmodifiable by Alice and unobservable by Bob.

The similarities and differences between classical and quantum information are summarized in Table 1.

Experimental studies of quantum information

The continuing maturation of the theory of quantum information and quantum computation has stimulated experimental work in a great variety of disciplines, in optics and quantum optics, in single-atom and single-ion research, and in several areas of precision spectroscopy. We will touch on some of the progress along these lines here. We will not mention here the very interesting prospects of using solid-state quantum technology—quantum dots, semiconductor microcavities, ultrasmall Josephson junctions, and so forth—to achieve quantum gate operation, which apparently lie several years further into the future.

'Flying qubits' will be needed to implement many of the quantum processing protocols described above. Because of developments in quantum cryptography, high-quality flying qubits in the form of photons travelling on optical fibres are now produced routinely in several laboratories. An important innovation, introduced by the Gisin group at the University of Geneva⁵⁶, helps to make reliable photon transmission through unreliable fibres a possibility. It involves the use of a Faraday mirror, which reflects any light that strikes it into an orthogonal polarization. In their scheme a strong coherent-state double pulse of light is sent from Bob to Alice on an

optical fibre; Alice attenuates it to one-photon intensity, sets the relative phase of the two pulses to obtain one of four quantum states of the photon, and finally Faraday-reflects this photon back to Bob. The Faraday reflection ensures that any distortions or variations of the propagating light mode due to birefringence (anisotropy of the index of refraction) in the photon transmission from Bob to Alice are undone in the return transmission. With this invention a remarkable interferometric stability is attained: the fringe visibility of their 23-km transmission system used as an interferometer has reached 99.98%, implying that the phase of the photon is reliable to 0.03 radians. This means that high-purity quantum states of light are being successfully transmitted in this system.

Their ability to store and process qubits with 'standing qubits' can greatly augment the capabilities of 'flying qubits' for the processing of quantum information. For example, the ability to do quantum computation in conjunction with quantum communication would qualitatively enhance the ability to do quantum cryptography, permitting the use of quantum repeater elements and opening up new techniques for defeating eavesdropping, as well as permitting cryptography over indefinitely long distances^{57,58}. With this in mind, workers have proposed a marriage of techniques from photon-fibre systems and trapped-atom (or ion) systems. In these schemes^{60,61}, a 'standing qubit' encoded in a state of the atom is mapped by an appropriate laser pulse⁵⁹ into the same qubit state of the photon state of a surrounding electromagnetic cavity, and can from there become a 'flying qubit' by leaking out into a propagating mode in free space or in an optical fibre. The unexpected feature of this procedure is the next step, in which the propagating photon then impinges on a replica of the sending system. If the photon wave packet has been tailored properly, this cavity-atom system can be made to recapture the photon into the atomic state by a suitable time-reverse of the sending procedure.

Although extensions of the original two-bit gate demonstration in cavity quantum electrodynamics⁶² have brought us closer to this goal of marrying flying and standing qubits we do not yet have an

elementary functioning prototype. Optical quantum electrodynamics (QED) experiments have not succeeded in entangling the states of two 'standing qubits', but such entanglement has been achieved in experiments in related areas, in microwave cavity QED (ref. 63) and in ion-trap studies⁶⁴.

Unfortunately, the controllable creation of entanglement with two-qubit quantum gates is only one of a formidable checklist of ingredients that a physical experiment must have if it is to realize a quantum computer. There are at least four other milestones which must be achieved⁶⁶: (1) The system should be extendible to a large number of qubits. (2) It must be possible to place the qubits reliably in the "0" or cleared state at the outset. (3) The decoherence rate must, as explained above, be very low (that is, below some suitable threshold). (4) It must be possible to do single-quantum sensitivity measurements (if only one copy of the quantum computer is available) or an accurate ensemble measurement in a qubit-specific fashion (if many copies of the quantum computer are available).

A full-scale experiment of any type to realize all of these criteria simultaneously is still a long way off. In the area of ion-trap research, concerted efforts are being undertaken by a number of experimental groups to realize the original Cirac and Zoller proposal⁶⁷ for ion-trap quantum computing, which created great excitement and interest five years ago. The proposal of these authors was nothing less than a scheme for realizing all of the requirements for quantum computation mentioned above: qubits are to be represented by the internal (spin) states of individual ions held in the electromagnetic trap; extending the number of qubits is to be achieved by adding more atoms to the trap. The techniques of laser cooling would serve to put the system in the "0" state. Coupling to the environment in the ion trap is low, and thus qubits with acceptable decoherence properties are known. The technique known as quantum-jump spectroscopy provides for the possibility of virtually single-quantum measurements of almost 100% efficiency. The heart of the proposal is a detailed scheme for the realization of two-qubit quantum operations: their procedure involves a coupling of the internal ion state with the quantum state of vibration of the ions in the trap. Because these oscillations involve collective modes of all the ions, entanglement of the internal ion states becomes possible. Unfortunately, one feature of the Cirac-Zoller computer—cooling to the ground state of motion of the trap—has proved to be very difficult to achieve, and this essential step has only been accomplished reliably by one group for one⁶⁵ or two⁶⁴ ions.

While it is clear that the trap ideas are on a steady track of progress, naturally workers in other fields hope that their techniques will enable them to leapfrog the atomic physicists and get ahead in the 'quantum computer race'. The proponents of nuclear magnetic resonance spectroscopy (NMR), as practised in organic chemistry, have made a bold move in this direction. NMR spectroscopy has many useful features for application to quantum computation: in a well-understood limit of rapidly tumbling molecules in solution, the hamiltonian of the nuclear spins of the molecule assumes a very simple form:

$$H = \sum_i \omega_i \sigma_z^i + \sum_{i,j} J_{ij} \sigma_z^i \sigma_z^j \quad (3)$$

(Here σ_z is the angular momentum operator of the spins, ω is the Zeeman splitting, and J is the exchange interaction parameter.) This depends only on the z -component of the nuclear spins (labelled i and j here). A system with this hamiltonian is well adapted to quantum computing^{68,69}: as it commutes with all σ_z^i operators, every computational basis state is an eigenstate. Thus, the state of the system only changes when a resonance pulse is applied, so that the dynamics of the system is entirely under external control. By appropriate frequency and time selectivity, an external pulse can perform a very finely tailored operation, for example, flipping one particular spin i if another specific spin j is up; this is the essence of the fundamental two-qubit XOR gate for quantum computing. In

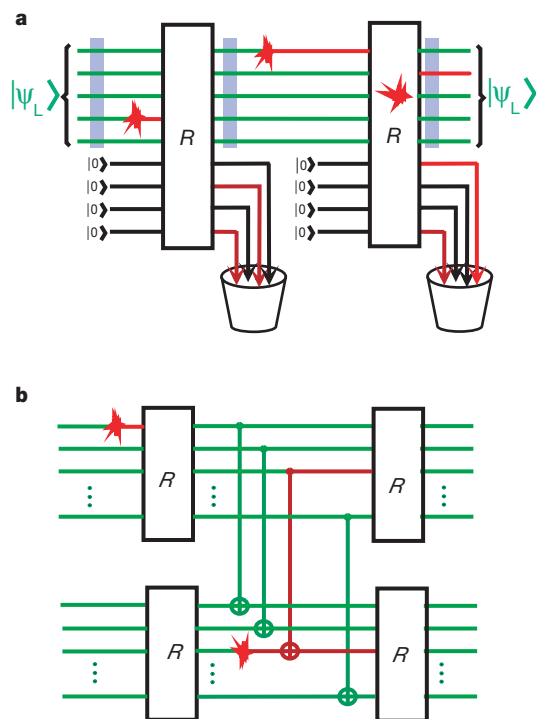


Figure 4 Fault-tolerant computation. **a**, Fault-tolerant error-correction circuit with cold ancillas coming in and corrupted ones being discarded. **b**, Performing the XOR operation on encoded data without decoding it.

addition, the pulse operations can be done much faster than the decoherence time of 1–10 seconds in a well-chosen molecule. Finally, in a situation in which many identical copies of the quantum computer are available (the many identical molecules in solution), the readout of the final result can be accomplished by an ensemble measurement of the transverse magnetization, a standard operation in NMR.

After the initial proposals of refs 70 and 71, there has been a flood of work on few-qubit systems, so numerous that we will just enumerate the accomplishments in this area briefly here: the action of two- and three-bit quantum gates has been demonstrated in the protons in 2,3-dibromothiophene and in 1-chloro-2-nitrobenzene⁷², the Deutsch–Jozsa algorithm⁷³ and the Grover algorithm⁷⁴ have been demonstrated using the H and ¹³C spins in chloroform, the three H and C spins in trichlorethylene have been used to simulate the synthesis of Greenberger–Horne–Zeilinger states⁷⁵, to perform teleportation³⁷ (see above), and to simulate the action of the three-qubit quantum error-correcting code⁷⁶ (the three C spins in alanine were also used in this last study), protons in cytosine have been used to implement the original Deutsch algorithm⁷⁷ as well as the quantum counting algorithm⁷⁸, and 2,3-dibromopropanoic acid has been used for some simple three-qubit gate arrays⁷⁹.

The NMR practitioners are pressing on to implement more quantum information processing in molecules with larger numbers of spins. However, there are a couple of large obstacles to immediately going on to large-scale quantum computation; probably these are not insuperable, but they may serve to make the progress of NMR quantum computing no faster than that in atomic physics or in other areas. One problem is just that the frequency-domain addressing which is used, in which each qubit has a distinct chemical shift ω_i , becomes difficult when the number of qubits grows large. A second problem (this will probably be the more immediate reason that the NMR technique will have to be radically modified to do quantum computation at a scale much greater than about 10 qubits) has to do with state preparation: the spin states of the molecules in the solution at room temperature are almost perfectly random, with a slight bias ϵ for the zero qubit state (typically ϵ , proportional to $k_B T$ divided by the nuclear Zeeman energy, is of the order of 10^{-6}). The number of molecules in the solution starting in the correct state, rather than the completely random state, scales with $\epsilon 2^{-n}$, where n is the number of spins in the molecule. The signal strength thus becomes exponentially small in the number of qubits, and all advantage gained from doing quantum computation is lost. This problem can be solved if ϵ can be increased to nearly one; there are innovative techniques in optical pumping which hold out the hope of doing this. While there is reason for optimism in these areas, we think that it will require many years of concerted effort.

We recall⁵⁹ the incident at a quantum computation meeting in Torino in 1995, when Shor offered a bet that the first factoring of a 500-digit number would be accomplished by a quantum and not a classical computer. There were no takers on the other side, but some commented that they would prefer to bet on a third possibility, that the Sun would burn up first. Although these sceptics have not been entirely silenced, on balance we are more in agreement with Shor than we were then. We think the odds in favour of the quantum computer have improved and will continue to increase slowly as more years of steady progress are made. □

1. Barenco, A. *et al.* Elementary gates for quantum computation. *Phys. Rev. A* **52**, 3457–3467 (1995).
2. Zurek, W. Decoherence and the transition from quantum to classical. *Phys. Today* **44**, 36–44 (1991).
3. Preskill, J. Reliable quantum computers. *Proc. R. Soc. Lond. A* **454**, 385–410 (1998).
4. Shor, P. W. in *Proceedings of the 35th Annual Symposium on the Foundations of Computer Science* 124–133 (IEEE Computer Society Press, Los Alamitos, California, 1994).
5. Ekert, A. & Jozsa, R. Shor's quantum algorithm for factorising numbers. *Rev. Mod. Phys.* **68**, 733–753 (1996).
6. Wiesner, S. Simulations of many-body quantum systems by a quantum computer. Preprint quant-ph/9603028 at (<http://xxx.lanl.gov>) (1996).

7. Abrams, D. S. & Lloyd, S. Simulations of many-body Fermi systems on a universal quantum computer. *Phys. Rev. Lett.* **79**, 2586–2589 (1997).
8. Grover, L. K. Quantum mechanics helps in searching for a needle in a haystack. *Phys. Rev. Lett.* **79**, 325–328 (1997).
9. Boyer, M., Brassard, G., Hoyer, P. & Tapp, A. Tight bounds on quantum searching. *Fortschr. Phys.* **46**, 493–506 (1998).
10. Ozhigov, Y. Quantum computers cannot speed up iterated applications of a black box. Preprint quant-ph/9712051 at (<http://xxx.lanl.gov>) (1997).
11. Terhal, B. M. *Quantum Algorithms and Quantum Entanglement*. Thesis, Univ. Amsterdam (1999).
12. Farhi, E., Goldstone, J., Gutmann, S. & Sipser, M. A limit on the speed of quantum computation in determining parity. *Phys. Rev. Lett.* **81**, 5442–5444 (1998).
13. Beals, R., Buhrman, H., Cleve, R., Mosca, M. & de Wolf, R. in *Proceedings of the 39th Annual Symposium on the Foundations of Computer Science* 352–361 (IEEE Computer Society Press, Los Alamitos, California, 1998).
14. Jozsa, R. & Schumacher, B. A new proof of the quantum noiseless coding theorem. *J. Mod. Opt.* **41**, 2343–2349 (1994).
15. Shor, P. W. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A* **52**, 2493–2496 (1995).
16. Calderbank, A. R. & Shor, P. W. Good quantum error correcting codes exist. *Phys. Rev. A* **54**, 1098–1105 (1996).
17. Steane, A. Multiple particle interference and quantum error correction. *Proc. R. Soc. Lond. A* **452**, 2551–2577 (1996).
18. Knill, E. & Laflamme, R. Theory of quantum error correcting codes. *Phys. Rev. A* **55**, 900–911 (1997).
19. Gottesman, D. A class of quantum error-correcting codes saturating the Hamming bound. *Phys. Rev. A* **54**, 1862–1868 (1996).
20. Bennett, C. H., DiVincenzo, D. P., Smolin, J. & Wootters, W. K. Mixed state entanglement and quantum error correction. *Phys. Rev. A* **54**, 3824–3851 (1996).
21. Helstrom, C. W. *Quantum Detection and Estimation Theory* (Academic, New York, 1976).
22. Kholevo, A. S. Bounds for the quantity of information transmitted by a quantum communication channel. *Problemy Peredachi Informatsii* **9**, 3–11 (1973); translated in *Problems Inf. Transmiss.* **9**, 177–183 (1973).
23. Holevo, A. S. Problems in the mathematical theory of quantum communication channels. *Rep. Math. Phys.* **12**, 273–278 (1977).
24. Schumacher, B., Westmoreland, M. & Wootters, W. K. Limitation on the amount of accessible information in a quantum channel. *Phys. Rev. Lett.* **76**, 3452–3455 (1997).
25. Holevo, A. S. The capacity of the quantum channel with general signal states. *IEEE Trans. Inf. Theory* **44**, 269–273 (1998).
26. Kholevo, A. S. Capacity of a quantum communications channel. *Problemy Peredachi Informatsii* **15**, 3–11 (1979); translated in *Problems Inf. Transmiss.* **15**, 247–253 (1979).
27. Sassaki, M., Kato, K., Izutsu, M. & Hirota, O. Quantum channels showing superadditivity in channel capacity. *Phys. Rev. A* **58**, 146–158 (1998).
28. Fuchs, C. A. Nonorthogonal quantum states maximize classical information capacity. *Phys. Rev. Lett.* **79**, 1162–1165 (1997).
29. Bennett, C. H. *et al.* Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels. *Phys. Rev. Lett.* **70**, 1895–1898 (1993).
30. Bennett, C. H. & Wiesner, S. J. Communication via one- and two-particle operators on Einstein–Podolsky–Rosen states. *Phys. Rev. Lett.* **69**, 2881–2884 (1992).
31. Bennett, C. H., Shor, P. W., Smolin, J. A. & Thapliyal, A. V. Entanglement enhanced classical capacity of noisy quantum channels. *Phys. Rev. Lett.* **83**, 3081–3084 (1999).
32. Mattle, K., Weinfurter, H., Kwiat, P. G. & Zeilinger, A. Dense coding in experimental quantum communication. *Phys. Rev. Lett.* **76**, 4656–4659 (1996).
33. Bouwmeester, D. *et al.* Experimental quantum teleportation. *Nature* **390**, 575–579 (1997).
34. Boschi, D., Branca, S., De Martini, F., Hardy, L. & Popescu, S. Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein–Podolsky–Rosen channels. *Phys. Rev. Lett.* **80**, 1121–1124 (1998).
35. Furusawa, A. *et al.* Unconditional quantum teleportation. *Science* **282**, 706–709 (1998).
36. Vaidman, L. Teleportation of quantum states. *Phys. Rev. A* **49**, 1473–1476 (1994).
37. Nielsen, M. A., Knill, E. & Laflamme, R. Complete quantum teleportation using nuclear magnetic resonance. *Nature* **396**, 52–55 (1998).
38. Cleve, R. & Buhrman, H. J. Substituting quantum entanglement for communication. *Phys. Rev. A* **56**, 1201–1204 (1997).
39. Buhrman, H., Cleve, R. & Wigderson, A. in *Proceedings of the 39th Annual ACM Symposium on the Theory of Computing* 63–68 (ACM Press, New York, 1998).
40. Bennett, C. H., Bernstein, H. J., Popescu, S. & Schumacher, B. Concentrating partial entanglement by local operators. *Phys. Rev. A* **53**, 2046–2052 (1996).
41. Lo, H.-K. & Popescu, S. Concentrating entanglement by local actions—beyond mean values. Preprint quant-ph/9707038 at (<http://xxx.lanl.gov>) (1997).
42. Vidal, G. Entanglement monotones. Preprint quant-ph/9807077 at (<http://xxx.lanl.gov>) (1998).
43. Linden, N., Popescu, S. & Sudbury, A. Non-local properties of multi-partite density matrices. *Phys. Rev. Lett.* **83**, 243–247 (1999).
44. Thapliyal, A. V. On multipartite pure-state entanglement. *Phys. Rev. A* **59**, 3336–3342 (1999).
45. Kempe, J. On multi-particle entanglement and its application to cryptography. *Phys. Rev. A* **60**, 910–916 (1999).
46. Horodecki, M., Horodecki, P. & Horodecki, R. Mixed state entanglement and distillation: is there a 'bound' entanglement in nature? *Phys. Rev. Lett.* **80**, 5239–5242 (1998).
47. Bennett, C. H. *et al.* Purification of noisy entanglement, and faithful teleportation via noisy channels. *Phys. Rev. Lett.* **76**, 722–725 (1996).
48. Deutsch, D. *et al.* Quantum privacy amplification and the security of quantum cryptography over noisy channels. *Phys. Rev. Lett.* **77**, 2818–2821 (1996); **80**, 2022 (1998) (errata).
49. Bennett, C. H. & Brassard, G. in *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India* 175–179 (IEEE, New York, 1984).
50. Mayers, D. Unconditional security in quantum cryptography. Preprint quant-ph/9802025 at (<http://xxx.lanl.gov>) (1998).
51. Lo, H.-K. & Chau, H. F. Unconditional security of quantum key distribution over arbitrarily long distances. *Science* **283**, 2050–2056 (1999).

52. Griffiths, R. B. & Niu, C.-S. Optimal eavesdropping in quantum cryptography. II. Quantum circuit. *Phys. Rev. A* **56**, 1173–1176 (1997).
53. Biham, E., Boyer, M., Brassard, G., van de Graaf, J. & Mor, T. Security of quantum key distribution against all collective attacks. *Phys. Rev. Lett.* **78**, 2256–2259 (1997).
54. Mayers, D. Unconditionally secure quantum bit commitment is impossible. *Phys. Rev. Lett.* **78**, 3414–3417 (1997).
55. Lo, H.-K. & Chau, H. F. Is quantum bit commitment really possible? *Phys. Rev. Lett.* **78**, 3410–3413 (1997).
56. Muller, A. *et al.* 'Plug and Play' systems for quantum cryptography. *Appl. Phys. Lett.* **70**, 793–795 (1997).
57. Briegel, H. J., Dür, W., Cirac, J. I. & Zoller, P. Quantum repeaters for communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
58. Dür, W., Briegel, H. J., Cirac, J. I. & Zoller, P. Quantum repeaters based on entanglement purification. *Phys. Rev. A* **59**, 169–181 (1999).
59. Bennett, C. H. & DiVincenzo, D. P. Quantum computing—towards an engineering era? *Nature* **377**, 389 (1995).
60. van Enk, S. J., Cirac, J. I. & Zoller, P. Ideal quantum communication over noisy channels: a quantum optical implementation. *Phys. Rev. Lett.* **78**, 4293–4296 (1997).
61. van Enk, S. J., Kimble, H. J., Cirac, J. I. & Zoller, P. Quantum communication with dark photons. *Phys. Rev. A* **59**, 2659–2664 (1999).
62. Mabuchi, H., Turchette, Q. A., Chapman, M. S. & Kimble, H. J. Real-time detection of individual atoms falling through a high-finesse optical cavity. *Opt. Lett.* **21**, 1393–1395 (1996).
63. Haroche, S., Brune, M. & Raimond, J. M. Experiments with single atoms in a cavity: entanglement, Schrödinger's cats and decoherence. *Phil. Trans. R. Soc. Lond. A* **355**, 2367–2380 (1997).
64. Turchette, Q. A. *et al.* Deterministic entanglement of two ions. *Phys. Rev. Lett.* **81**, 3631–3634 (1998).
65. Monroe, C., Meekhof, D. M., King, B. E., Itano, W. M. & Wineland, D. J. Demonstration of a fundamental quantum logic gate. *Phys. Rev. Lett.* **75**, 4714–4717 (1995).
66. DiVincenzo, D. P. & Loss, D. Quantum information is physical. *Superlatt. Microstruct.* **23**, 419–432 (1998).
67. Cirac, J. I. & Zoller, P. Quantum computations with cold trapped ions. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
68. Lloyd, S. A potentially realizable quantum computer. *Science* **261**, 1569–1571 (1993).
69. Lloyd, S. Envisioning a quantum supercomputer. *Science* **263**, 695 (1994).
70. Cory, D. G., Fahmy, A. F. & Havel, T. F. Ensemble quantum computing by nuclear magnetic resonance spectroscopy. *Proc. Natl. Acad. Sci. USA* **94**, 1634–1639 (1997).
71. Gershenfeld, N. A. & Chuang, I. L. Bulk spin resonance quantum computation. *Science* **275**, 350–356 (1997).
72. Cory, D. G., Price, M. D. & Havel, T. F. Nuclear magnetic resonance spectroscopy: an experimentally accessible paradigm for quantum computing. *Physica D* **120**, 82–101 (1998).
73. Chuang, I. L., Vandersypen, L. M. K., Xinlan Zhou, Leung, D. W. & Lloyd, S. Experimental realization of a quantum algorithm. *Nature* **393**, 143–146 (1998).
74. Chuang, I. L., Gershenfeld, N. & Kubinec, M. Experimental implementation of fast quantum searching. *Phys. Rev. Lett.* **80**, 3408–3411 (1998).
75. Laflamme, R., Knille, E., Zurek, W. H., Catasti, P. & Mariappan, S. V. S. NMR Greenberger Horne Zeilinger states. *Phil. Trans. R. Soc. Lond. A* **356**, 1941–1948 (1998).
76. Cory, D. G. *et al.* Experimental quantum error correction. *Phys. Rev. Lett.* **81**, 2152–2155 (1998).
77. Jones, J. A. & Mosca, M. Implementation of a quantum algorithm on a nuclear magnetic resonance quantum computer. *J. Chem. Phys.* **109**, 1648–1653 (1998).
78. Jones, J. A. & Mosca, M. Approximate quantum counting on an NMR ensemble quantum computer. *Phys. Rev. Lett.* **83**, 1050–1053 (1999).
79. Linden, N., Barjat, H. & Freeman, R. An implementation of the Deutsch Jozsa algorithm on a three qubit NMR quantum computer. *Chem. Phys. Lett.* **296**, 61–67 (1998).

Acknowledgements

This work was supported by the US Army Research office.

Correspondence and requests for materials should be addressed to C.H.B. (e-mail: bennetc@watson.ibm.com).



Experimental entanglement of four particles

C. A. Sackett*, D. Kielpinski*, B. E. King†‡, C. Langer*, V. Meyer*, C. J. Myatt*, M. Rowe*, Q. A. Turchette‡, W. M. Itano*, D. J. Wineland* & C. Monroe*

* Time and Frequency Division, National Institute of Standards and Technology, Boulder, Colorado 80303, USA

† Atomic Physics Division, NIST, Gaithersburg, Maryland 20899, USA

‡ Research Electro-Optics, 1855 S. 57th Ct., Boulder, Colorado 80301, USA

Quantum mechanics allows for many-particle wavefunctions that cannot be factorized into a product of single-particle wavefunctions, even when the constituent particles are entirely distinct. Such ‘entangled’ states explicitly demonstrate the non-local character of quantum theory¹, having potential applications in high-precision spectroscopy², quantum communication, cryptography and computation³. In general, the more particles that can be entangled, the more clearly nonclassical effects are exhibited^{4,5}—and the more useful the states are for quantum applications. Here we implement a recently proposed entanglement technique⁶ to generate entangled states of two and four trapped ions. Coupling between the ions is provided through their collective motional degrees of freedom, but actual motional excitation is minimized. Entanglement is achieved using a single laser pulse, and the method can in principle be applied to any number of ions.

Most experimental demonstrations of entanglement to date have relied on the selection of data from random processes, such as the preparation and detection of photon pairs in parametric down-conversion^{7–9} or of atoms in a thermal beam¹⁰. All methods of this type suffer from inescapable signal degradation when entanglement of larger numbers of particles is attempted, as the probability of randomly generating the appropriate conditions decreases exponentially. For instance, in the experiment of ref. 7, two-photon entangled states could be generated and detected at a rate of roughly 1,000 per second, three-photon states at a rate of 30 per hour, and four-photon states at an extrapolated rate of several per year. Trapped ions have been suggested as a system in which such effects might be avoided¹¹, and we have demonstrated two-particle entanglement in a deterministic way¹². By ‘deterministic’ we mean that the desired state could be produced with a high degree of certainty at a user-specified time¹³, which is necessary for avoiding the degradation described above. However, that experiment relied on the particular behaviour of two ions in a quadrupole radio-frequency trap, and could not easily be applied to larger numbers of particles.

The entanglement technique proposed by Mølmer and Sørensen^{6,14} can be understood by considering a pair of spin-half charged particles confined together in a harmonic potential. The energy levels of this system are illustrated in Fig. 1, where $\hbar\omega_0$ is the internal energy splitting of each particle, and ν is the oscillation frequency of a particular collective mode of the particles in the trap. Using laser-cooling and optical-pumping techniques¹⁵, the particles are initially prepared in their spin-down internal state and in the ground state of their collective motion: $|\psi\rangle = |\downarrow\downarrow 0\rangle$. By applying optical fields oscillating at $\omega_0 + \nu - \delta$ and $\omega_0 - \nu + \delta$, the two-step transition from $|\downarrow\downarrow 0\rangle$ to $|\uparrow\uparrow 0\rangle$ is driven. For sufficiently large δ , the intermediate states $|\uparrow\downarrow 1\rangle$ and $|\downarrow\uparrow 1\rangle$ are negligibly occupied, so that no motional excitation occurs. The resulting interaction hamiltonian, in the rotating-wave approximation and the Lamb–Dicke limit, is then

$$H = \frac{\hbar\tilde{\Omega}}{2}(|\uparrow\uparrow\rangle\langle\downarrow\downarrow| + |\downarrow\downarrow\rangle\langle\uparrow\uparrow| + |\uparrow\downarrow\rangle\langle\downarrow\uparrow| + |\downarrow\uparrow\rangle\langle\uparrow\downarrow|) \quad (1)$$

with $\tilde{\Omega} = \eta^2\Omega^2/\delta$ when the single particle $|\downarrow\rangle \leftrightarrow |\uparrow\rangle$ transition has

Rabi frequency Ω and the Lamb–Dicke parameter is η . For an excitation involving momentum transfer $\hbar k$ and a total particle mass of M , η is given by $(\hbar k^2/2M\nu)^{1/2}$. Entanglement is achieved by applying H for a time $t = \pi/2\tilde{\Omega}$, making the spin wavefunction $|\psi_s\rangle = (|\uparrow\uparrow\rangle - i|\downarrow\downarrow\rangle)/\sqrt{2}$. This spin state is in fact created for any initial motional state $|n\rangle$, so long as the Lamb–Dicke criterion $\eta^2 n \ll 1$ is satisfied.

In order for the intermediate states $|\uparrow\downarrow 1\rangle$ and $|\downarrow\uparrow 1\rangle$ to be negligibly occupied, the detuning δ must be large compared to the transition linewidth $\eta\Omega$. However, it is clear from the expression for $\tilde{\Omega}$ that the entanglement speed is maximized for small δ , and in fact the technique can still be applied for $\delta \approx \eta\Omega$ (refs 14, 16). Although motional excitation does then occur to some degree, for select values of δ the excitation vanishes at precisely the time that the entangled spin state is created. The condition for this to occur is

$$\delta/\eta\Omega = 2\sqrt{m} \quad (2)$$

for any integer m , and the maximum excitation during the pulse then has mean quantum number $\bar{n} = 1/2m$. Our experiment is operated with $m = 1$.

As discussed in ref. 6, the entanglement method is scalable in the sense that precisely the same operation can be used to generate the N -particle entangled state

$$|\psi_N\rangle = (|\uparrow\uparrow\ldots\uparrow\rangle + i^{N+1}|\downarrow\downarrow\ldots\downarrow\rangle)/\sqrt{2} \quad (3)$$

if N is any even number, while for N odd, $|\psi_N\rangle$ can be generated using one entanglement pulse accompanied by a separate independent rotation of each particle’s spin.

If the ions are uniformly illuminated, the Mølmer and Sørensen scheme^{6,14} requires that they all participate equally in the intermediate motional excitation, which implies that the only suitable mode for arbitrary N is the centre-of-mass mode. However, this mode has a practical disadvantage because in our experiments fluctuating ambient electric fields cause it to heat at a significant rate. Although for large δ the entanglement operation is largely independent of the motion, so that heating is unimportant, in the small- δ case it is necessary that motional decoherence be avoided. Modes involving only relative ion motion couple to higher moments of the field, so heating of these modes is negligible¹⁵. For $N = 2$ and $N = 4$, such modes do exist in which each particle participates with equal amplitude¹⁷. In both cases, they are symmetric ‘stretch’ modes, in which alternating ions oscillate out of

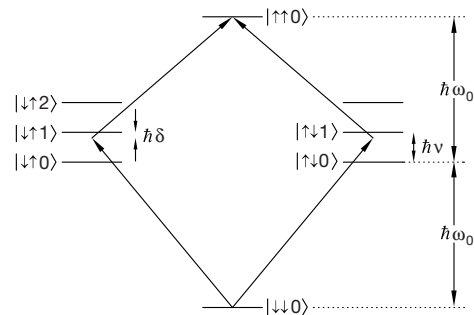


Figure 1 Entanglement scheme for two particles. Each ion is initially prepared in the $|\downarrow\downarrow\rangle$ internal state, and the collective motion of the pair is cooled to its ground state $|0\rangle$. Laser fields oscillating near $\omega_0 + \nu$ and $\omega_0 - \nu$ couple the $|\downarrow\downarrow\rangle$ and $|\uparrow\uparrow\rangle$ states as shown. By detuning the single transition frequencies by a small amount δ , the populations of the $|\uparrow\downarrow 1\rangle$ and $|\downarrow\uparrow 1\rangle$ states are kept small. Then, by driving the double transition for the appropriate time, the entangled state $(|\uparrow\uparrow\rangle - i|\downarrow\downarrow\rangle)/\sqrt{2}$ is created. For four ions, the same procedure generates the state $(|\uparrow\uparrow\uparrow\uparrow\rangle + i|\downarrow\downarrow\downarrow\downarrow\rangle)/\sqrt{2}$. We note that in the actual experiment, each of the single transitions shown is itself a two-photon Raman transition, driven by a pair of laser beams; the entire process therefore consists of a four-photon transition.

phase. We use these modes here. Excitation of the centre-of-mass mode still affects the experiment, as the motion in spectator modes modifies the coupling strength to the mode of interest^{14,15}. For this reason, we initially cool both the centre-of-mass and stretch modes to near their ground state.

The experiment was performed using $^9\text{Be}^+$ ions confined in a miniature linear radio-frequency trap¹⁸, with the N ions lying in a line along the trap's weak axis. Two spectrally resolved ground-state hyperfine levels comprise the effective spin-half system, with $|\downarrow\rangle \equiv |F=2, m_F=-2\rangle$ and $|\uparrow\rangle \equiv |F=1, m_F=-1\rangle$, where F is the total angular momentum quantum number, and $\hbar m_F$ is the projection of the angular momentum along the quantization axis defined by an externally applied magnetic field. The hyperfine splitting $\omega_0/2\pi$ is approximately 1.25 GHz. Coherent coupling between $|\downarrow\rangle$ and $|\uparrow\rangle$ is provided by stimulated Raman transitions. The two Raman laser beams have a wavelength of 313 nm with a difference frequency near ω_0 , and are perpendicular, with their difference wave-vector lying along the line of ions. They are detuned ~ 80 GHz blue of the $2P_{1/2}$ excited state, with intensities giving $\Omega/2\pi \approx 500$ kHz. For both the two- and four-ion experiments, the desired stretch-mode frequency $\nu/2\pi$ was 8.8 MHz, giving $\eta = 0.23/N^{1/2}$. The two driving frequencies required for the entanglement operation are generated by frequency-modulating one of the Raman beams using an electro-optic modulator.

After the entanglement operation, the ions are probed by illuminating them with a circularly polarized laser beam tuned to the $2S_{1/2}(F=2, m_F=-2) \leftrightarrow 2P_{3/2}(F=3, m_F=-3)$ cycling transition. Each ion in $|\downarrow\rangle$ fluoresces brightly, leading to the detection of ~ 15 photons per ion on a photomultiplier tube during a 200- μs detection period. In contrast, an ion in $|\uparrow\rangle$ remains nearly dark. Because the number of photons detected from a spin-down ion fluctuates according to Poisson statistics, in a single experiment the number of spin-down ions can be determined with only a limited accuracy. For the data reported, each experiment was repeated 1,000 times under the same conditions, and the resulting photon-number distribution fitted to a sum of poissonians to determine the probabilities P_j for j ions to be in $|\downarrow\rangle$. The results are given in Table 1, and show that in both cases, the probabilities for all N ions to be in the same state are large compared to the probabilities for the other cases. This is characteristic of the states $|\psi_N\rangle$, although the fact

Table 1 Characterization of two-ion and four-ion states

N	P_0	P_1	P_2	P_3	P_4	$\rho_{\uparrow\uparrow}$
2	0.43	0.11	0.46	—	—	0.385
4	0.35	0.10	0.10	0.10	0.35	0.215

N is the number of ions, P_j denotes the probability that j ions were measured to be in $|\downarrow\rangle$, and $\rho_{\uparrow\uparrow}$ denotes the amplitude of the density matrix element $\rho_{|\uparrow\ldots\uparrow\rangle, |\uparrow\ldots\uparrow\rangle}$. Uncertainties in $\rho_{\uparrow\uparrow}$ and the $N=2$ populations are ± 0.01 , and uncertainties in the $N=4$ populations are ± 0.02 .

that the middle probabilities are non-zero indicates that we do not generate the entangled states with perfect accuracy.

In order to prove that we are generating a reasonable approximation to $|\psi_N\rangle$, it is necessary to prove that the populations of $|\uparrow\ldots\uparrow\rangle$ and $|\downarrow\ldots\downarrow\rangle$ are coherent. In terms of the density matrix for the system, ρ , we must measure the far off-diagonal element $\rho_{|\uparrow\ldots\uparrow\rangle, |\downarrow\ldots\downarrow\rangle}$, whose amplitude will be abbreviated $\rho_{\uparrow\downarrow}$. This can be achieved by applying a simple analysis pulse to the ions before observing them. If the Raman difference frequency is set to ω_0 (and the frequency modulator turned off), each ion i undergoes ordinary Rabi oscillations, evolving according to the hamiltonian

$$H_i = \frac{\hbar\Omega}{2}(e^{i\phi}|\uparrow\rangle\langle\downarrow|_i + e^{-i\phi}|\downarrow\rangle\langle\uparrow|_i) \quad (4)$$

where ϕ is the phase of the difference frequency relative to that of the entanglement pulse. This hamiltonian is applied for time $\pi/2\Omega$ (a $\pi/2$ pulse), and the parity

$$\Pi \equiv \sum_{j=0}^N (-1)^j P_j \quad (5)$$

is observed while ϕ is varied. As seen in Fig. 2, for N ions Π oscillates as $\cos N\phi$, and the amplitude of this oscillation is in fact $2\rho_{\uparrow\downarrow}$ (ref. 2). The resulting values are given in Table 1. From the data shown in the table, our state preparation fidelity

$$F \equiv \langle\psi_N|\rho|\psi_N\rangle = \frac{1}{2}(P_{\uparrow\uparrow} + P_{\downarrow\downarrow}) + \rho_{\uparrow\downarrow} \quad (6)$$

can be determined, where $P_{\uparrow\uparrow}$ is the population of $|\uparrow\ldots\uparrow\rangle$ and $P_{\downarrow\downarrow}$ is the population of $|\downarrow\ldots\downarrow\rangle$. For $N=2$ we achieve $F = 0.83 \pm 0.01$, while for $N=4$, $F = 0.57 \pm 0.02$.

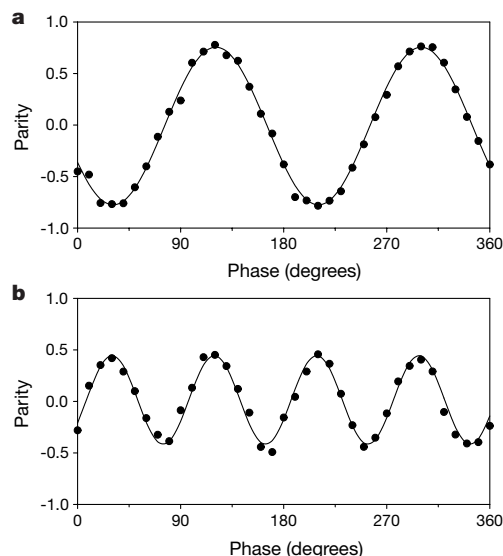


Figure 2 Determination of $\rho_{\uparrow\downarrow}$. **a**, Interference signal for two ions; **b**, four ions. After the entanglement operation of Fig. 1, an analysis pulse with relative phase ϕ is applied on the single-ion $|\downarrow\rangle \leftrightarrow |\uparrow\rangle$ transition. As ϕ is varied, the parity of the N ions oscillates as

$\cos N\phi$, and the amplitude of the oscillation is twice the magnitude of the density-matrix element $\rho_{\uparrow\downarrow}$. Each data point represents an average of 1,000 experiments, corresponding to a total integration time of roughly 10 s for each graph.

The fact that $\rho_{(11)}$ is non-zero is still insufficient to guarantee entanglement. To be explicit, a system with density matrix ρ exhibits N -particle entanglement only if no decomposition $\rho = \sum_k p_k |\psi_k\rangle\langle\psi_k|$ exists with all the $|\psi_k\rangle$ factorizable into products of wavefunctions that depend on fewer than N particles. For example, if $|\psi\rangle_i = (|\uparrow\rangle_i + |\downarrow\rangle_i)/\sqrt{2}$ is a state of ion i , then the four-particle state $|\psi\rangle_1|\psi\rangle_2|\psi\rangle_3|\psi\rangle_4$ is not entangled, but still has $\rho_{(11)} = 1/16$. We note that these are the types of states studied in liquid-state nuclear magnetic resonance experiments¹⁹. Alternatively, for $|\psi\rangle_{12} = (|\uparrow\rangle_1|\uparrow\rangle_2 + |\downarrow\rangle_1|\downarrow\rangle_2)/\sqrt{2}$ and $|\psi\rangle_{34} = (|\uparrow\rangle_3|\uparrow\rangle_4 + |\downarrow\rangle_3|\downarrow\rangle_4)/\sqrt{2}$, the state $|\psi\rangle_{12}|\psi\rangle_{34}$ exhibits two-particle, but not four-particle entanglement, and has $\rho_{(11)} = 1/4$.

To establish that we are actually observing N -particle entanglement, consider an arbitrary factorizable wavefunction

$$|\psi_F\rangle = [a|\uparrow \dots \uparrow\rangle_X + b|\downarrow \dots \downarrow\rangle_X + \dots][c|\uparrow \dots \uparrow\rangle_Y + d|\downarrow \dots \downarrow\rangle_Y + \dots] \quad (7)$$

where X and Y refer to two distinct subsets of the N particles, with $|\uparrow \dots \uparrow\rangle_X$ indicating the state with all particles in subset X spin-up, and similarly for the other terms. Normalization of the factor wavefunctions requires $|a|^2 + |b|^2 \leq 1$ and $|c|^2 + |d|^2 \leq 1$, which can be combined and rewritten as

$$(|a| - |c|)^2 + 2|ac| + (|b| - |d|)^2 + 2|bd| \leq 2 \quad (8)$$

Since the squared terms on the left are positive, equation (8) implies that $|ac| + |bd| \leq 1$, and in turn that $(|ac| + |bd|)^2 \leq 1$. Expanding the square yields the desired relation²⁰:

$$P_{(1)} + P_{(1)} + 2\rho_{(11)} = 2F \leq 1 \quad (9)$$

where $P_{(1)} = |ac|^2$, $P_{(1)} = |bd|^2$, and $\rho_{(11)} = |abcd|$ are the previously defined quantities. Since equation (9) holds for any separable wavefunction, it must also hold for any separable density matrix. Both our $N = 2$ and $N = 4$ experiments give $F > 1/2$, so the states they produce exhibit N -particle entanglement.

Quantifying the amount of entanglement present is a more difficult question. A variety of measures of entanglement have been proposed, but most are difficult to calculate even numerically^{21,22}. For $N = 2$, there is an explicit formula for the ‘entanglement of formation’ E as a function of ρ (ref. 23). Although we have not reconstructed the entire two-particle density matrix, the populations measured place sufficient bounds on the unmeasured elements to determine that $E \approx 0.5$. This indicates that roughly two pairs of our ions would be required to carry the same quantum information as a single perfectly entangled pair.

In the four-ion case, no explicit formula for entanglement is known. The data do indicate that our density matrix can be expressed approximately as

$$\rho = 0.43|\psi_4\rangle\langle\psi_4| + 0.57\rho_{\text{incoh}} \quad (10)$$

where $|\psi_4\rangle$ is the desired state and ρ_{incoh} is completely incoherent (that is, diagonal). These coefficients are determined directly from the value of $\rho_{(11)}$ in Table 1, together with the fact that no evidence for other off-diagonal matrix elements was observed. To determine a measure of entanglement, however, it is necessary to decompose ρ as a sum of $|\psi_4\rangle$ and a ‘worst-case’ factorizable matrix ρ_F , which can be accomplished as

$$\rho = 0.13|\psi_4\rangle\langle\psi_4| + 0.87\rho_F. \quad (11)$$

Note that equations (10) and (11) both describe the same physical state, but that in equation (11), ρ_F consists of a specific mixture of two- and three-particle entangled states that is highly unlikely to occur in our experiments. In either description, it is clear that our state-preparation accuracy is limited.

The source of decoherence in our experiments is not entirely clear, but evidence suggests that it is related to intensity fluctuations in the Raman laser beams, a problem we are working to understand

and correct²⁴. The presence of decoherence, and the fact that it affects the four-ion experiment more strongly than the two-ion one, illustrates the need to carefully define the sense in which our entanglement operation is ‘scalable’. Any entanglement experiment is more sensitive to decoherence as the number of particles involved is increased, unless sufficient accuracy can be achieved for error-correction schemes to be usefully applied. Such schemes are thought to require an error rate of the order of 10^{-4} per operation²⁵, and we are certainly far from this regime. However, even if such a level of fidelity were to be achieved, applications such as quantum computing still require very large entangled states to be generated in a reasonable amount of time and using a reasonable amount of resources. The method demonstrated here is important in this regard, since it uses only a single operation and requires a time that scales roughly as $N^{1/2}$.

In the language of quantum information science, we have realized a four-quantum-bit logic gate. This system is relevant for the future development of quantum information technology, as such states may be used to implement quantum error-detection schemes²⁶ or to make rudimentary demonstrations of quantum algorithms^{27,28}. Entanglement of four particles is also interesting in its own right, as such states can show strong violations of local realism⁵. Even the two-particle Bell’s inequality measurement would be interesting to implement, as the near-perfect detection efficiency for ions would eliminate the ‘fair sampling’ hypothesis which has been required in other experiments²⁹. In addition to improved fidelity, applications such as these do require the ability to perform individual manipulation and detection of each ion, but this is not expected to be a severe experimental challenge³⁰. $\square$

Received 11 February; accepted 15 February 2000.

1. Bell, J. *Speakable and Unsayable in Quantum Mechanics* (Cambridge Univ. Press, 1987).
2. Bollinger, J., Itano, W. M., Wineland, D. & Heinzen, D. Optimal frequency measurements with maximally correlated states. *Phys. Rev. A* **54**, R4649–R4652 (1996).
3. Lo, H.-K., Popescu, S. & Spiller, T. (eds) *Introduction to Quantum Computation and Information* (World Scientific, Singapore, 1997).
4. Pan, J.-W., Bouwmeester, D., Daniell, M., Weinfurter, H. & Zeilinger, A. Experimental test of quantum non-locality in three-photon Greenberger–Horne–Zeilinger entanglement. *Nature* **403**, 515–519 (2000).
5. Mermin, N. D. Extreme quantum entanglement in a superposition of macroscopically distinct states. *Phys. Rev. Lett.* **65**, 1838–1840 (1990).
6. Molmer, K. & Sørensen, A. Multiparticle entanglement of hot trapped ions. *Phys. Rev. Lett.* **82**, 1835–1838 (1999).
7. Bouwmeester, D., Pan, J.-W., Daniell, M., Weinfurter, H. & Zeilinger, A. Observation of three-photon Greenberger–Horne–Zeilinger entanglement. *Phys. Rev. Lett.* **82**, 1345–1349 (1999).
8. Kwiat, P. et al. New high-intensity source of polarization-entangled photon pairs. *Phys. Rev. Lett.* **75**, 4337–4341 (1995).
9. Tittle, W., Brendel, J., Zbinden, H. & Gisin, N. Violation of Bell inequalities by photons more than 10 km apart. *Phys. Rev. Lett.* **81**, 3565–3566 (1998).
10. Hagley, E. et al. Generation of Einstein–Podolsky–Rosen pairs of atoms. *Phys. Rev. Lett.* **79**, 1–5 (1997).
11. Cirac, J. & Zoller, P. Quantum computations with cold trapped ions. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
12. Turchette, Q. et al. Deterministic entanglement of two trapped ions. *Phys. Rev. Lett.* **81**, 3631–3634 (1998).
13. Law, C. & Kimble, J. Deterministic generation of a bit-stream of single-photon pulses. *J. Mod. Opt.* **44**, 2067–2074 (1997).
14. Molmer, K. & Sørensen, A. Entanglement and quantum computation with ions in thermal motion. Pre-print quant-ph/0002024 at (xxx.lanl.gov) (2000).
15. King, B. et al. Cooling the collective motion of trapped ions to initialize a quantum register. *Phys. Rev. Lett.* **81**, 1525–1528 (1998).
16. Milburn, G. Simulating nonlinear spin models in an ion trap. Pre-print quant-ph/9908037 at (xxx.lanl.gov) (1999).
17. James, D. Quantum dynamics of cold trapped ions with applications to quantum computation. *Appl. Phys. B* **66**, 181–190 (1998).
18. Turchette, Q. et al. Heating of trapped ions from the quantum ground state. Pre-print quant-ph/0002040 at (xxx.lanl.gov) (2000).
19. Fitzgerald, R. What really gives a quantum computer its power? *Phys. Today* **53**, 20–22 (2000).
20. Bennett, C. et al. Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.* **76**, 722–725 (1996).
21. Vedral, V., Plenio, M., Rippin, M. & Knight, P. Quantifying entanglement. *Phys. Rev. Lett.* **78**, 2275–2279 (1997).
22. Lewenstein, M. & Sanpera, A. Separability and entanglement of composite quantum systems. *Phys. Rev. Lett.* **80**, 2261–2264 (1998).
23. Wootters, W. K. Entanglement of formation of an arbitrary state of two qubits. *Phys. Rev. Lett.* **80**, 2245–2248 (1998).
24. Wineland, D. et al. Experimental issues in coherent quantum-state manipulation of trapped atomic ions. *J. Res. NIST* **103**, 259–328 (1998).

25. Preskill, J. Battling decoherence: the fault tolerant quantum computer. *Phys. Today* **52**, 24–30 (1999).
26. Vaidman, L., Goldenberg, L. & Wiesner, S. Error prevention scheme with four particles. *Phys. Rev. A* **54**, R1745–R1748 (1996).
27. Shor, P. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.* **26**, 1484–1509 (1997).
28. Grover, L. Quantum mechanics helps in searching for a needle in a haystack. *Phys. Rev. Lett.* **79**, 325–328 (1997).
29. Eberhard, P. Background level and counter efficiencies required for a loophole-free Einstein–Podolsky–Rosen experiment. *Phys. Rev. A* **47**, R747–R750 (1993).
30. Nägerl, H. *et al.* Laser addressing of individual ions in a linear ion trap. *Phys. Rev. A* **60**, 145–148 (1999).

Acknowledgements

We thank J. Bollinger and E. A. Cornell for comments on the manuscript. This work was supported by the US National Security Agency, Army Research Office and Office of Naval Research.

Correspondence and requests for materials should be addressed to C. M.
(e-mail: monroe@boulder.nist.gov).

Quantum distribution of protons in solid molecular hydrogen at megabar pressures

Hikaru Kitamura^{*†‡}, Shinji Tsuneyuki^{*}, Tadashi Ogitsu^{*§}
& Takashi Miyake^{||}

^{*} Institute for Solid State Physics, University of Tokyo, Roppongi, Tokyo 106-8666, Japan

[†] Department of Space Physics and Astronomy, Rice University, Houston, Texas 77251-1892, USA

[‡] Applied Theoretical and Computational Physics Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

[§] National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

^{||} Joint Research Center for Atom Technology, Tsukuba, Ibaraki 305-0046, Japan

Solid hydrogen, a simple system consisting only of protons and electrons, exhibits a variety of structural phase transitions at high pressures. Experimental studies¹ based on static compression up to about 230 GPa revealed three relevant phases of solid molecular hydrogen: phase I (high-temperature, low-pressure phase), phase II (low-temperature phase) and phase III (high-pressure phase). Spectroscopic data suggest that symmetry breaking, possibly related to orientational ordering^{1,2}, accompanies the transition into phases II and III. The boundaries dividing the three phases exhibit a strong isotope effect³, indicating that the quantum-mechanical properties of hydrogen nuclei are important. Here we report the quantum distributions of protons in the three phases of solid hydrogen, obtained by a first-principles path-integral molecular dynamics method. We show that quantum fluctuations of protons effectively hinder molecular rotation—that is, a quantum localization occurs. The obtained crystal structures have entirely different symmetries from those predicted by the conventional simulations which treat protons classically.

The structures of these broken-symmetry phases have been extensively investigated both by experimental^{1–3} and by theoretical^{4–9} studies, though the results are still controversial. In most theoretical studies, even when they precisely compare total energies of model structures on the basis of electronic-structure calculations, quantum fluctuations of protons are usually neglected, or roughly discussed, perhaps due to some technical reason. But as shown by our previous study on impurity muonium (a bound state comprising

a positive muon and an electron) and hydrogen in crystalline silicon¹⁰, quantum states of light particles may exhibit distributions significantly different from those expected from stability analyses based on the classical potential-energy surface. It is therefore desirable to examine quantitatively the quantum-mechanical properties of protons in solid hydrogen, especially when we are concerned with its symmetry breaking under high pressure. Natoli *et al.*¹¹ treated both electrons and protons with quantum Monte Carlo methods; their simulations were, however, performed with trial wavefunctions based on the fixed static configuration of the protons, and were restricted to the molecules centred on the hexagonal close packed (h.c.p.) lattice sites.

The first-principles path-integral molecular dynamics (FP-PIMD) method enables us to incorporate quantum-mechanical properties of protons in conventional Car–Parrinello-type first-principles simulations. The basic formalism was first presented by Marx and Parrinello¹². Recently this scheme was further developed by the present authors¹⁰, and we use this in the present study. We consider a supercell containing N ($= 64$) atoms, which is subject to the periodic boundary conditions. To represent the quantum properties of protons in the path-integral formalism, imaginary time $\beta\hbar$ (where β is the inverse temperature in units of Boltzmann constant) is divided into P finite time slices; each proton is thus represented by a polymer consisting of P ‘beads’ interacting via intrapolymeric harmonic forces¹³. Exchange effects between protons are neglected in this study. Assuming the Born–Oppenheimer approximation, interatomic forces between protons at each atomic configuration and at each time slice are determined by electronic-structure calculation based on the density-functional theory (DFT). The resultant classical system of NP particles is simulated by molecular dynamics (MD) at constant volume and temperature with the aid of the Nosé–Hoover chain thermostat¹⁴. In the evaluation of the exchange-correlation potential in the DFT calculation, we adopt the generalized gradient approximation (GGA)¹⁵. The interaction between a proton and an electron has been described by the norm-conserving non-local pseudopotential of the Troullier–Martins type¹⁶; the electronic wavefunctions have been expanded in plane waves with a cut-off energy of 50 Ry. Integration over the first Brillouin zone has been achieved by sampling 8 uniform k -points. Energy eigenvalues of the lowest 32 occupied bands have been explicitly calculated, since the system should remain an insulator in the relevant pressure range¹. The ground-state electron density has been obtained by minimizing the total energy functional through the conjugated-gradient method¹⁷. We have verified, through static DFT calculations with these parameter conditions, that the potential energy surface associated with molecular rotation in the $Pca2_1$ structure obtained earlier by

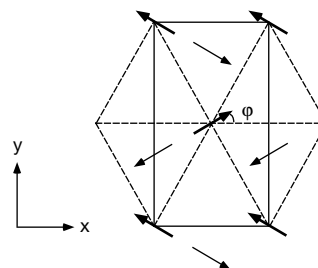


Figure 1 Illustration of the $Pca2_1$ structure projected onto the x – y plane. Thick arrows represent molecules on the α -plane and thin arrows depict those on the β -plane, both pointing towards the positive- z hemisphere. These two planes are apart from each other by $c/2$ in the z -direction. The quantity θ measures the angle between a molecular axis and the z -axis of the crystal: φ is the corresponding azimuthal angle. The $Cmc2_1$ structure is obtained by setting $\varphi = 90^\circ$.

the linearized augmented plane-wave calculation⁴ can be reproduced with reasonable accuracy.

We have performed FP-PIMD simulations for three combinations of densities and temperatures corresponding to the three phases³: phase I (120 GPa, 300 K), phase II (130 GPa, 80 K), and phase III (180 GPa, 100 K). The corresponding cell size has been estimated by the experimental equation of state extrapolated to pressures above 120 GPa (ref. 18). The values of P have been chosen as 32 for phase I and 64 for phases II and III, which we have confirmed to be large enough to reproduce the distribution function associated with the intramolecular vibration of an isolated molecule at corresponding temperatures.

As the DFT calculations for different imaginary time slices are mutually independent, they can be performed efficiently through parallel computation. For each time slice, the band-structure calculation can also be performed using parallel computing with respect to the band index. The simulations have been performed by using 128 processing elements on the Hitachi SR 2201 at the Computer Center of the University of Tokyo and 512 processing elements on the CP-PACS computer at the Research Center for Computational Physics of the University of Tsukuba; both computers have a theoretical peak speed of 0.3×10^9 floating-point operations per second per processing element. The CPU time is about 72 seconds for each imaginary time slice per conjugated-gradient step when using eight processing elements. About ten conjugated-gradient steps have been necessary per each MD step. Though the relative stability of the h.c.p.-based structure and the cubic-close-packed structure is competitive, previous static calculations¹⁹ predicted that the former is more stable under megabar pressures. Each run has therefore been started with the atomic configuration corresponding to $Pca2_1$ (Fig. 1). The total number of MD steps achieved after thermalization is 2,100, 1,500 and 1,800 for phases I, II and III, respectively, with a corresponding time step of 0.1 fs, 0.15 fs and 0.1 fs.

Figure 2 displays the isosurface of the quantum distributions of protons thus obtained for the three phases; structural differences associated with symmetry breakings are clearly observed. In phase I, we find that the distribution is almost spherical and shows no clear orientational order. The bond centre of each molecule lies on the h.c.p. lattice site. These results are consistent with experimental observations^{1,3} that rotational disorder persists in the entire regime of phase I up to 150 GPa. In phase II, on the other hand, the

distribution is predicted to be no longer spherical; its projection onto the x - y plane exhibits anisotropy with enhanced distribution along the y -axis of the crystal, indicating that the crystalline structure is $Cmc2_1$. The bond centre of each molecule does not deviate significantly from the h.c.p. lattice site. In phase III, the rotational order is even more conspicuous, showing that molecular rotation is strongly prohibited. Here, molecular bond centres form the $Cmca$ structure, which belongs to a space group different from h.c.p.; the molecules originally located on the β -plane of the h.c.p. lattice deviate from the lattice sites by 0.51 Å towards the positive- y direction, while the molecules on the α -planes are fixed. We note that the distribution in phase I is not perfectly spherical; further MD samplings are necessary to clarify this anisotropy more quantitatively. The structural difference observed among the three phases is, however, remarkable, and we believe that the basic features of these distributions would remain unchanged even if we were to obtain further simulation data.

Biermann *et al.*⁵ recently performed FP-PIMD simulations of solid hydrogen at 50 K and at pressures above 150 GPa, finding a “string type” structure²⁰, which we did not observe in the present simulations. We consider that the results we report here are the more reliable for the following reasons. First, Biermann *et al.* adopted the local-density approximation (LDA) for the exchange correlation potential, which gives poor results for some systems, including hydrogen. More importantly these workers adopted a single k -point (Γ -point) sampling in the Brillouin-zone integration. It has been shown⁴ that sampling of many k points is required to obtain a convergent result on the potential energy surface and especially its angular dependence relevant to orientational orderings.

We plot in the upper panels of Fig. 3 the distributions of bond angles (see the caption of Fig. 1 for definition) for molecules on the α -planes. The upper three panels represent the FP-PIMD results for the three phases. In phase I, though a slight non-uniformity exists, the distribution is almost uniform on the θ - φ plane, indicating rotational disorder. In phase II, the distribution has maxima at $(\theta, \varphi) \approx (50^\circ, 90^\circ)$ and at its equivalent point $(130^\circ, 270^\circ)$, which corresponds to $Cmc2_1$ structure as we have seen in Fig. 2. Despite the rotational ordering, we find that the values of θ and φ are distributed over fairly wide ranges around the maxima. On the other hand, in phase III, the distribution is strongly localized at around $(\theta, \varphi) \approx (70^\circ, 90^\circ)$; the molecules now behave like classical librions

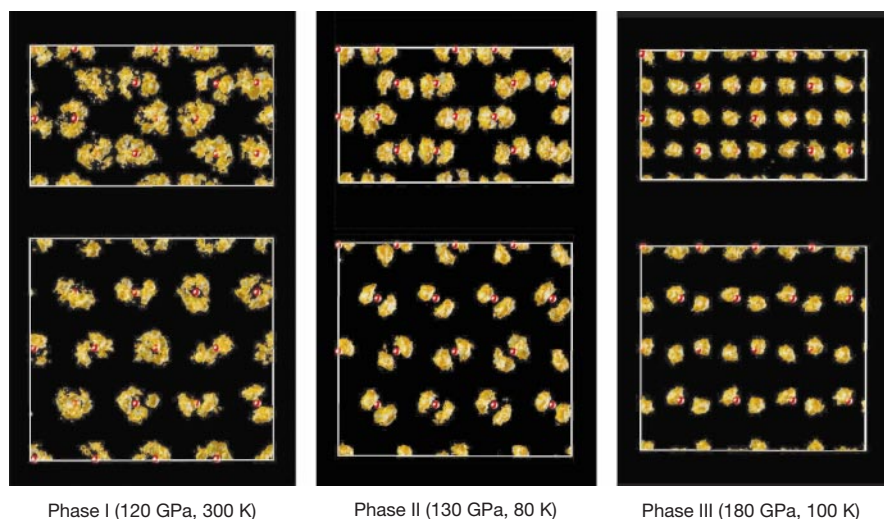


Figure 2 Quantum distribution of protons in the phases I, II and III of solid hydrogen obtained by the FP-PIMD method. The isosurface corresponding to 1/4 of the maximum density is plotted. The h.c.p. lattice sites are shown by red balls. For each phase, the upper plot represents a projection onto the y - x plane and the lower plot is that onto the

y - z plane, where the horizontal axes coincide with the y -axis in both plots. The cell size (in Å) in the (x, y, z) directions is (3.711, 6.428, 5.882), (3.669, 6.355, 5.808) and (3.509, 6.078, 5.527) for phase I, II and III, respectively.

in which full rotation is strongly hindered. (In this connection, we note that qualitative differences between the experimental roton/libron spectra for phases II and III have been previously pointed out²: the roton bands in phase III are virtually independent of pressure, whereas they are replaced by strongly pressure-dependent libron spectra in phase III.)

For comparison, we have also performed simulations with $P = 1$ so that protons are treated as classical particles. The resultant angular distributions are displayed in the lower panels of Fig. 3. The distribution for phase I is completely uniform, and seems qualitatively similar to that obtained by FP-PIMD. We note that in phase II, the distribution is almost indistinguishable from that for phase I and virtually no orientational ordering can be observed, contrary to the prediction by FP-PIMD. In phase II, the angular distribution seems qualitatively the same as that obtained by FP-PIMD, but the amount of deviation of molecular bond centres on the β -plane from their h.c.p. lattice points turns out to be only about 0.05 Å, which is 10% of that observed in the FP-PIMD results. The structure is therefore off-centred $Cmc2_1$, not $Cmca$.

We usually expect the distribution of a particle to become spatially more delocalized when quantum zero-point motion is taken into account. But what is observed here in phase II is an apparently opposite effect, which may be called ‘quantum localization’ (or ‘quantum confinement’); the same trend is also observed in phase I, though not as distinctly as in phase II. Quantum localization can occur in a multi-dimensional potential field when the particle motion in a certain direction is restricted by confinement in its orthogonal direction²¹. For illustration, consider a particle of mass M in the two-dimensional potential $V(x, y) = V_0[(x/a)^2 - 1]^2 + M\omega^2(|x|)y^2/2$ with $V_0 > 0$, $a > 0$ and $\omega(|x|)$ being positive and an increasing function of $|x|$. If $\hbar\omega(|a|)/2 \gg \hbar\omega(0)/2 + V_0$, a quantum particle is localized at $x = 0$ rather than at the potential minima $x = \pm a$, because the latter are unstable as a result of the large zero-point energy associated with the confinement in the y -direction, whereas a classical particle prefers thermal hopping between the two potential minima $x = \pm a$. This mechanism offers an explanation for the previous finding that muonium in silicon is confined at the centre of a silicon cage because of the quantum effect¹⁰. We expect that the same mechanism may apply to the rotational potential $V(\theta, \varphi)$ in hydrogen, though a detailed analysis will be required. It is unlikely that the quantum localization is an artefact owing to the lack of MD samplings, as we have confirmed that the distribution obtained by classical MD for the first 1,000 steps looks qualitatively similar to that obtained for the last 1,000 steps.

Kohanoff *et al.*⁷ performed constant-pressure (110–150 GPa) *ab initio* MD for phase II with the GGA, where protons were regarded as classical and quantum corrections due to zero-point vibrations

Table 1 Relative stability of the orientationally ordered structures

Structures	120 GPa			180 GPa		
	<i>Pca2</i> ₁	<i>Cmc2</i> ₁	<i>Cmca</i>	<i>Pca2</i> ₁	<i>Cmc2</i> ₁	<i>Cmca</i>
r (Å)	0.735	0.736	0.778	0.734	0.740	0.780
θ, φ (deg)	62, 50	46, 90	66, 90	62, 49	61, 90	70, 90
E_{static} (meV per atom)	0	1.4	24.5	0	2.7	−6.8
E_{vibron} (meV/atom)	0	−5.4	−27.2	0	−12.2	−27.2
E_{total} (meV/atom)	0	−4.0	−2.7	0	−9.5	−34.0
ω_{Raman} (cm ^{−1})	3,570	3,260	2,800	3,310	2,770	2,260
ω_{exp} (cm ^{−1})		4,130			3,840	

DFT-GGA calculations of the static total energies E_{static} , zero-point energies of vibrons E_{vibron} and their sum $E_{\text{total}} = E_{\text{static}} + E_{\text{vibron}}$ (relative to the values for *Pca2*₁ structure) are listed for various structures. For each structure, E_{static} has been determined so that it is minimized with respect to the bond length r and two angles θ and φ ; for *Cmc2*₁ and *Cmca* structures φ has been fixed at 90° by definition. We have estimated the vibron zero-point energies E_{vibron} in an approximate way as follows. The static total energy $E_{\text{static}}(r)$ of the system, calculated as a function of r , is fitted to the following functional form (Morse potential):

$$E_{\text{static}}(r) = D \left[\exp\left(-\frac{2(r-r_m)}{R}\right) - 2 \exp\left(-\frac{r-r_m}{R}\right) \right]$$

where r_m is the optimum bond length, and D and R are fitting parameters. We use $E_{\text{vibron}} = \hbar\omega_{\text{HM}}/2 - \hbar^2\omega_{\text{HM}}^2/16D$, where $\omega_{\text{HM}} = \sqrt{2D/\mu R^2}$ appearing in the first term is the harmonic frequency with μ denoting the reduced mass of the hydrogen molecule; the second term represents anharmonicity. The vibron frequencies $\omega_{\text{Raman}} = \omega_{\text{HM}} - \hbar/\mu R^2$ for Raman-active modes and the corresponding experimental data¹ ω_{exp} are also shown.

were added separately in the harmonic approximation: they found the *Pca2*₁ structure to be the most stable at zero temperature. It remains to be seen whether their simulation would predict such rotational order at higher temperatures relevant to the present study.

In order to interpret these results in terms of relative stability among various lattice structures, we have performed static DFT-GGA calculations and have evaluated static total energies E_{static} for optimized *Pca2*₁, *Cmc2*₁ and *Cmca* structures. The results are summarized in Table 1. As far as the static energies are concerned, the *Pca2*₁ structure is more stable than *Cmc2*₁ by a small amount (about 1.4–2.7 meV per atom). The *Cmca* structure is quite unstable at 120 GPa, but becomes energetically favourable at higher pressures. The largest contribution to the zero-point motion of protons arises from vibrons¹. As shown in Table 1, the vibron zero-point energies E_{vibron} are smallest in *Cmca* and largest in *Pca2*₁. As a result, at 120 GPa, the sum $E_{\text{static}} + E_{\text{vibron}}$ is smallest in *Cmc2*₁, consistent with the FP-PIMD results for phase II. At 180 GPa, the smallness of the vibron zero-point energy acts to stabilize *Cmca* relative to other on-centred h.c.p.-based structures which do not accompany deviation of bond centres. This explains why the present FP-PIMD simulation in phase III predicts *Cmca* but classical simulation does not, and also accounts for the

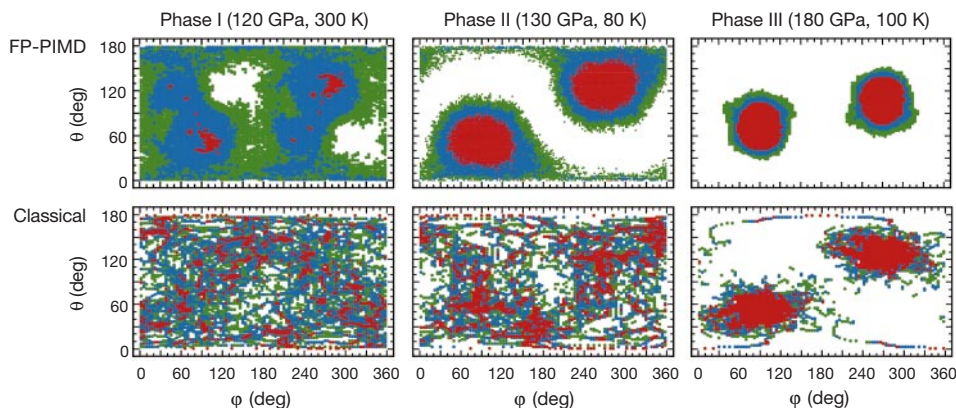


Figure 3 Distribution function $g(\theta, \varphi)$ of bond angles. The upper panels show the results of the FP-PIMD calculations; the lower panels, those of the classical simulations. The red, blue and green points indicate the regions where the distribution function satisfies

$g(\theta, \varphi) > 2$, $1 < g(\theta, \varphi) \leq 2$, and $0.5 < g(\theta, \varphi) \leq 1$, respectively. The function g is normalized so that $(1/4\pi) \int_0^\pi \int_0^{2\pi} d\theta d\varphi g(\theta, \varphi) = 1$.

experimental fact³ that the pressure value at the II–III phase boundary is lower for H₂ than for D₂.

There exist uncertainties in our predictions of the vibron frequencies, owing to their sensitive dependence on the treatment of the exchange-correlation potentials. As anharmonicity becomes more and more important at higher pressures^{1,22}, the frequency of Raman-active vibrons becomes a steeply decreasing function of pressure. Though this is qualitatively consistent with the experimentally observed “vibron softening” (ref. 1), quantitatively the frequencies predicted by the present theory are too small to explain the experimental data¹, as shown in Table 1. When the LDA is adopted, the resultant frequencies are smaller than the GGA predictions (by about 400 cm⁻¹ at 180 GPa) and agreement with experimental data gets worse. Therefore, both the LDA and the GGA seem unsatisfactory for reproducing the vibron frequencies accurately at high pressures. As Table 1 indicates, since E_{vibron} depends on the lattice structures more sensitively than E_{static} does, a better calculation beyond the GGA may predict a structure other than *Cmca* at 180 GPa. Moreover, though we have ignored the conduction bands, a full band calculation based on the LDA or GGA predicts hydrogen to be a metal in the *Cmca* structure at 180 GPa (ref. 9), which is inconsistent with both the experimental evidence and our assumption. This is also likely to arise from the failure of the DFT-GGA scheme. The significance of the work we report here is that, within the DFT-GGA formalism, we have successfully taken into account proton quantum effects on the basis of *ab initio* path-integral formalism to show that the distributions of protons in the broken-symmetry phases look entirely different from those predicted by classical simulations. □

Received 2 September 1999; accepted 13 January 2000.

- Mao, H.-k. & Hemley, R. J. Ultrahigh-pressure transitions in solid hydrogen. *Rev. Mod. Phys.* **66**, 671–692 (1994).
- Mazin, I. I., Hemley, R. J., Goncharov, A. F., Hanfland, M. & Mao, H.-k. Quantum and classical orientational orderings in solid hydrogen. *Phys. Rev. Lett.* **78**, 1066–1069 (1997).
- Hemley, R. J. & Mao, H.-k. in *Elementary Processes in Dense Plasmas* (eds Ichimaru, S. & Ogata, S.) 271–282 (Addison-Wesley, Reading, 1995).
- Mazin, I. I. & Cohen, R. E. Insulator-metal transition in solid hydrogen: Implication of electronic-structure calculations for recent experiments. *Phys. Rev. B* **52**, R8597–R8600 (1995).
- Biermann, S., Hohl, D. & Marx, D. Proton quantum effects in high pressure hydrogen. *J. Low Temp. Phys.* **110**, 97–102 (1998); Quantum effects in solid hydrogen at ultra-high pressure. *Solid State Commun.* **108**, 337–341 (1998).
- Edwards, B. & Ashcroft, N. W. Spontaneous polarization in dense hydrogen. *Nature* **388**, 652–655 (1997).
- Kohano, J., Scandolo, S., Chiarotti, G. L. & Tosatti, E. Solid molecular hydrogen: The broken symmetry phase. *Phys. Rev. Lett.* **78**, 2783–2786 (1997).
- Nagao, K. & Nagara, H. Theoretical study of Raman and infrared active vibrational modes in highly compressed solid hydrogen. *Phys. Rev. Lett.* **80**, 548–551 (1998).
- Nagao, K., Takezawa, T. & Nagara, H. *Ab initio* calculation of optical-mode frequencies in compressed solid hydrogen. *Phys. Rev. B* **59**, 13741–13753 (1999).
- Miyake, T., Ogitsu, T. & Tsuneyuki, S. Quantum distributions of muonium and hydrogen in crystalline silicon. *Phys. Rev. Lett.* **81**, 1873–1876 (1998).
- Natoli, V., Martin, R. M. & Ceperley, D. Crystal structure of molecular hydrogen at high pressure. *Phys. Rev. Lett.* **74**, 1601–1604 (1995).
- Marx, D. & Parrinello, M. *Ab initio* path-integral molecular dynamics. *Z. Phys. B* **95**, 143–144 (1994); *Ab initio* path integral molecular dynamics: Basic ideas. *J. Chem. Phys.* **104**, 4077–4082 (1996).
- Chandler, D. & Wolynes, P. G. Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids. *J. Chem. Phys.* **74**, 4078–4095 (1981).
- Tuckerman, M. E., Marx, D., Klein, M. L. & Parrinello, M. Efficient and general algorithms for path integral Car-Parrinello molecular dynamics. *J. Chem. Phys.* **104**, 5579–5588 (1996).
- Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996); Errata **78**, 1396 (1997).
- Troullier, N. & Martins, J. L. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B* **43**, 1993–2005 (1991).
- Payne, M. C., Teter, M. P., Allen, D. C., Arias, T. A. & Joannopoulos, J. D. Iterative minimization techniques for *ab initio* total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.* **64**, 1045–1097 (1992).
- Loubeyre, P. et al. X-ray diffraction and equation of state of hydrogen at megabar pressures. *Nature* **383**, 702–704 (1996).
- Nagara, H. & Nakamura, T. Stable phases of solid hydrogen at megabar pressures and at zero temperature. *Phys. Rev. Lett.* **68**, 2468–2471 (1992).
- Hohl, D., Natoli, V., Ceperley, D. M. & Martin, R. M. Molecular dynamics in dense hydrogen. *Phys. Rev. Lett.* **71**, 541–544 (1993).
- Miyake, T., Kusakabe, K. & Tsuneyuki, S. Inverse isotope effects in the surface diffusion of atoms. *Surf. Sci.* **363**, 403–408 (1996).
- Ashcroft, N. W. Pairing instabilities in dense hydrogen. *Phys. Rev. B* **41**, 10963–10971 (1990).

Acknowledgements

We thank H. Nagara for discussions. H.K. was supported by the Japan Society for the Promotion of Science for Young Scientists. This work was partially supported by the Research for the Future Program of the Japan Society for the Promotion of Science, and by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture, Japan.

Correspondence and requests for materials should be addressed to S.T. (e-mail: stsune@issp.u-tokyo.ac.jp).

Towards the clarity limit in optical fibre

Gordon A. Thomas*†, Boris I. Shraiman*, Paul F. Glodis‡ & Michael J. Stephen§

* Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974-0636, USA

‡ Network Products, Lucent Technologies, Norcross, Georgia 30071-2992, USA

§ Rutgers, State University of New Jersey, New Brunswick, New Jersey 08854-8019, USA

† Present address: George R. Harrison Spectroscopy Laboratory, Room 6-008, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

An important scientific and technological goal^{1–4} in the field of optical communications is the achievement of the clarity limit in optical fibres—that is, ensuring that the SiO₂ glass from which fibres are made is as transparent as possible. The clarity of the wavelength transmission window (and the width of that window) in existing fibres is already sufficient to form the basis of a world-wide optical communication system^{3,4}, yet it is still limited by contamination of the fibre by water. Here we measure the spatial distribution of water in the glass rods from which optical fibres are drawn and explain the distribution quantitatively with a mathematical model of diffusion^{5,6} in a medium with essentially perfect cylindrical symmetry. Our analysis shows that the water enters from the outside of the rod and diffuses into the molten, flowing glass much faster than is expected from extrapolation of low-temperature measurements^{6,7}. Our elucidation

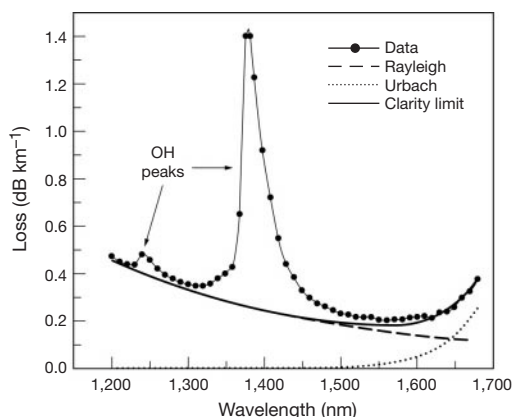


Figure 1 Loss of light intensity in an optical fibre as a function of the wavelength of the light. The solid points with a solid line interpolated through them are data from a fibre drawn from part of the glass rod measured in Figs 2, 3 and 4. The thick solid line is obtained from the sum of the behaviour calculated from Rayleigh scattering (dashed line) and the behaviour from the Urbach tail of the SiO₂ atomic vibrational spectrum (dotted line).

of the physics underlying the contamination process has already led to the fabrication of dry fibres^{8,9}, which have a clarified and broadened communications window. The improved operational range of wavelengths should yield applications for new lasers, optical amplifiers and detectors, and should substantially increase the information-carrying capacity of optical communications systems.

The effect of water on the clarity of optical fibres is illustrated in Fig. 1. The solid circles with a line interpolated between them are our measurement of the loss in the intensity of light transmitted through a fibre drawn from one rod we studied, as a function of the wavelength of the light. We measured this loss by the conventional method^{3,4} of comparing the intensity of the light transmitted through two relatively long fibres of different lengths. The two peaks in Fig. 1 result from the water and arise from absorption by the first overtones of the set of five partially overlapping OH vibrational modes¹⁰. These peaks ride on top of a curve (the thick solid line in Fig. 1) that may be the clarity limit for SiO₂ glass. This background curve is the sum of the scattering and absorption from the SiO₂ molecules themselves. The scattering is described by the Rayleigh theory^{3,4} and varies as the inverse of the wavelength to the fourth power, as shown by the dashed curve in Fig. 1, and the absorption is an exponential Urbach tail^{3,4}, shown by the dotted curve.

We used an unconventional method, interrupting one of the standard processes³ for manufacturing optical fibre at a key step and studying the condition of the glass rod at that point. (There are other processes for making fibre¹⁻⁴ that differ from the one we have studied in important ways. Our results are applicable to these processes indirectly in that we determine the high temperature diffusion of water in any fabrication method). Our sample preparation began with an SiO₂ tube about 1 m in length, heated

with a torch 10 cm in length that burned H₂ with O₂ as it moved slowly along the length of the tube. The thermal symmetry of the tube is improved by using a semi-cylindrical torch and by rotating the tube. We deposited layers of ultra-pure glass on its inside surface, for the core of the fibre. We then collapsed the tube at a temperature of 2,300 °C over a period of three hours, with a given section of the tube hot for one-tenth of that time because of the size of the torch. Finally, we removed a slice of the rod before resuming the process and drawing the rod into a fibre.

We cooled the rod, freezing in the water that had entered, cut and polished a slice 2 cm in thickness, and probed the slice with a beam of light 0.5 mm in diameter. Although the slice was clear to the eye, we observed the absorption of an infrared light beam by the OH fundamental vibrations¹¹ as shown in Fig. 2. We first compared the spectrum of light transmitted through the glass to that through a vacuum and found that there was no observable OH absorption over a region 3 mm in radius around the centre of the deposited layers. We concluded that this region was dry and so we normalized the transmitted intensity $I(r)$ at various radial distances r from the centre (as labelled in Fig. 2), to the intensity at $r = 0$, that is, $I(0)$. The effective absorption coefficient is plotted, as defined by $\alpha = \ln[I(0)/I(r)]/d$, where $d (= 2 \text{ cm})$ is the thickness of the slice of glass. (Our analysis does not depend on the normalization, but the display of the OH absorption mode has the most nearly flat background with the method shown.) Figure 2a shows the spectra taken within the starting glass tube and Fig. 2b shows spectra within some of the deposited layers. In the deposited layers, we find that the OH peak is smaller and so we have expanded the scale in Fig. 2b. The curves show that the smallest r at which we can reliably discern the OH peak is $r = 3.5 \text{ mm}$. The background spectrum is not flat because of variations in the composition of the deposited glass¹⁻⁴ that are used to guide the communications signal through the core of the fibre.

The absorption coefficient in Fig. 2 is linearly proportional to the concentration of OH per SiO₂ molecule c_{OH} using the calibration¹¹: c_{OH} (in p.p.m.) = α (in cm⁻¹)/759 for the peak of the OH absorption, at a wavelength of 2.72 micrometres. We measure the integrated area under the absorption curves relative to the background and scale the values from the curve at $r = 7 \text{ mm}$ where the background is negligible. The relative values of concentration in OH units per million SiO₂ units are plotted in Fig. 3 as solid circles.

This flow of water (or more precisely, the OH impurity site) into the glass (which also flows) changes the concentration c with time t and radial distance according to $dc/dt + \mathbf{v} \cdot \nabla c = D \nabla^2 c$. Here, the

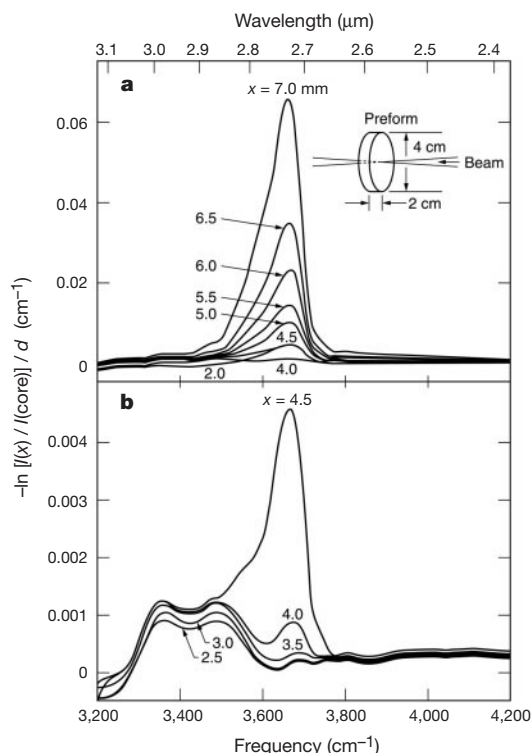


Figure 2 Absorption coefficient of transmitted light as a function of the energy of the light, used to measure the concentration of OH in our slice of glass rod. **a**, **b**, Each of the curves is obtained from the absorption of a narrow beam of light at a position along a radius r of the rod, a distance from the centre (core) as labelled. The peak at and near a wavelength of 2.72 micrometres (**a**) arises from the absorption by the set of overlapping fundamental vibrational modes of the OH molecules¹⁰.

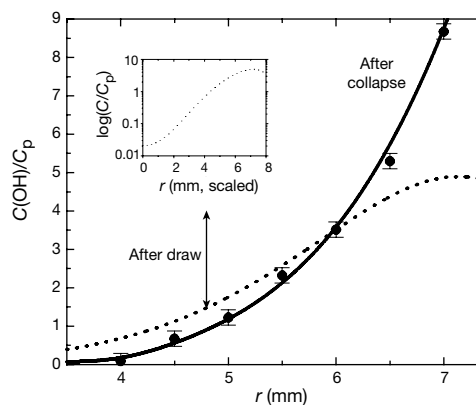


Figure 3 Measured and calculated relative values of OH concentration, $C(\text{OH})/C_p$ (see text), as a function of position in the glass rod. The data (solid circles with error bars) are obtained from Fig. 2 and the solid curves are calculated for OH diffusion with glass flow at a temperature of 2,300 °C for a time of 0.3 hours. The dotted line is the distribution calculated in a fibre after it is drawn from the same rod and re-scaled to the radius of the rod. Inset, the same curve over a larger spatial range with the concentration on a log scale.

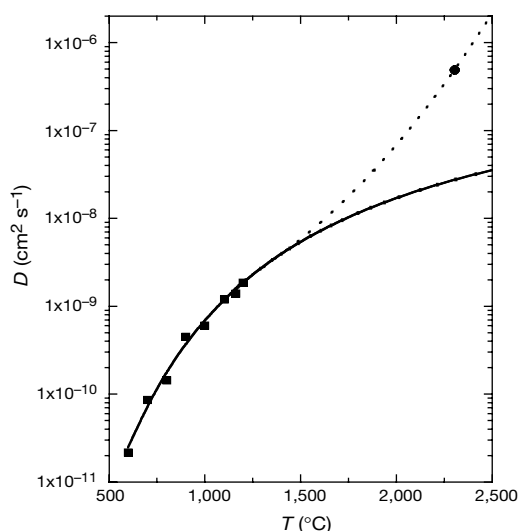


Figure 4 Measured values of the diffusion coefficient D as a function of temperature T . The solid line is an exponential fit to the low T data (solid squares). The dotted line is an

interpolation between the low temperature data and the value measured here (solid circle), showing the substantial enhancement in D .

diffusion coefficient is D , $\mathbf{v}$ is the velocity of flow of the glass given by $-\hat{r}(dR_0/dt)R/r$ and the concentration on the outer surface is fixed at c_0 . The Peclet number is $Pe = \nu R/D$, which for our case is about 150, indicating that the glass flow dominates the OH flux. The characteristic length of the flow is $r_D = (DL/U)^{1/2}$, where U is the average flow speed $\langle dR_0/dt \rangle$ of the outer radius and L is the rod length. In the case of slow diffusion (r_D small compared to the radius of the rod R) and no flow of the glass, the diffusion profile is given by $c(r) = c_0(R/r)^{1/2}[1 - \text{erf}\{(R-r)/(2r_D)\}]$. This form does not provide a good fit to our spatial profile of OH concentrations. We include the flow of the glass in the diffusion equation and solve numerically using lagrangian coordinates to obtain the solid curve in Fig. 3, which we fit to the data with the coefficient $D = 0.5 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$, with a statistical uncertainty of 10%. The contribution to the absorption coefficient from the initial amount of water in the substrate tube is negligible in our 2-cm-thick disk.

Using this value of D , we can also calculate the spatial distribution of OH in the glass after it is drawn into a fibre. The large velocities of the glass motion in both the radial and longitudinal directions must be included, because the length of the rod increases from 1 m to 75 km, and $U \approx 700 \text{ cm s}^{-1}$. Again we solve numerically using moving lagrangian coordinates and obtain the result shown as the dotted line in Fig. 3 (re-scaled to the size of the glass rod). The inset in Fig. 3 shows the OH distribution in fibre on a logarithmic scale with the spatial scale extending into the centre of the fibre. This curve shows that the measured distribution of OH in the glass rod produces a significant concentration of OH in the core of the fibre, as also indicated by the value $r_D \approx 10 \text{ }\mu\text{m}$, which is comparable to the radius of the glass rod after the draw. We measured the loss directly in the fibre, as shown in Fig. 1, and found an amount consistent with our estimate.

We can compare the value of D , determined from the fit in Fig. 3, with previous measurements⁷ as shown in Fig. 4. Our value (solid circle) is considerably above the extrapolation (dashed line on the semi-log plot) of the previous measurements (solid squares). Our fit to the low temperature data gives the form $D_{LT} = D_{LTO} \exp(-E_{LT}/kT)$, with $D_{LTO} = 10^{-6} \text{ cm}^2 \text{ s}^{-1}$ and $E_{LT} = 0.8 \text{ eV}$, or, at $T = 2,300 \text{ }^\circ\text{C}$, $D_{LT} = 2.8 \times 10^{-8} \text{ cm}^2 \text{ s}^{-1}$. The enhancement of our measured D over this extrapolated value results from a reduction in the activation energy (or barrier height) E for OH hopping above the glass-to-liquid transition T_g . This temperature depends on the thermal history of the glass⁵⁻⁷, but it occurs below $1,800 \text{ }^\circ\text{C}$ and well below our experimental $T = 2,300 \text{ }^\circ\text{C}$. We expect a decreased E above T_g

because the viscosity shows a decreased E above T_g in glasses generally⁵⁻⁷ and because the Stokes–Einstein relationship⁵⁻⁷ implies that a similar decrease in E should occur in both the diffusion and the viscosity. If we attribute the enhanced D entirely to a change in E , then E decreases by a factor of four from 0.8 eV below T_g to 0.2 eV at $T = 2,300 \text{ }^\circ\text{C}$. We speculate that this large reduction in E results from reductions in the strengths of the inter-atomic potentials above the melting temperature. Because both of these effects are substantial, we suggest that the large reduction in E is reasonable.

Our results give an improved understanding of the diffusion of the water in the flowing glass during the existing manufacturing process for optical fibre. An important part of devising a new process will thus be simply to reduce or eliminate the water that currently comes from the $\text{O}_2\text{--H}_2$ torch. On the basis of a preliminary version of this work, we published a patent⁸ describing a manufacturing process using a dry heat source, such as an O_2 plasma torch, but otherwise the same as the process described above. Subsequently, dry fibre has been produced⁹ and is now commercially available with a loss curve closely approximating the curve in Fig. 1 that may be the clarity limit. □

Received 14 December 1999; accepted 10 February 2000.

1. Keck, D. B., Maurer, R. D. & Schultz, P. C. On the ultimate lower limit of attenuation in glass optical waveguides. *Appl. Phys. Lett.* **22**, 307–309 (1973).
2. Keck, D. B., Schultz, P. C. & Zimar, F. Method of forming optical waveguide fibers. (US Patent No. 3737292, 1973).
3. MacChesney, J. B. & DiGiovanni, D. J. Materials development of optical fiber. *J. Am. Ceram. Soc.* **73**, 3537–3556 (1990).
4. Keiser, G. *Optical Fiber Communications* 1–55 (McGraw-Hill, New York, 1983).
5. Limoge, Y. in *Diffusion in Materials* (ed. Lasker, A. L.) 601–631 (NATO Advanced Study Institute Series, Kluwer, Boston, 1990).
6. Doremus, R. H. *Glass Science* (Wiley, New York, 1994).
7. Molson, A. J. & Roberts, J. P. Water in silica glass. *Proc. Chem. Soc. Lond.* **59**, 388–399 (1960).
8. Partus, F., Thomas, G. A., Atkins, R. & Fleming, J. Optical fiber with low OH impurity and communication system using the optical fiber. (US Patent No. 5692087, 1995).
9. Chang, K. & Kalish, D. New hydrogen aging loss mechanisms in the 1400nm window. *Proc. Opt. Fiber Commun. Conf. PD22*, 1–3 (1999).
10. Stone, J. & Walrafen, G. E. Overtone vibrations of OH groups in fused silica optical fibers. *J. Chem. Phys.* **76**, 1712–1722 (1982).
11. Shelby, J. E., Vitko, J. & Benner, R. E. Quantitative determination of the hydroxyl content of vitreous silica. *J. Am. Ceram. Soc.* **60**, C59–C60 (1982).

Acknowledgements

We acknowledge valuable assistance from P. LeMaire, F. Partus, R. Atkins, J. Fleming, K. Chang and D. Kalish.

Correspondence and requests for materials should be addressed to G. A. T. (e-mail: gordonat@mit.edu).

Direct oxidation of hydrocarbons in a solid-oxide fuel cell

Seungdoo Park, John M. Vohs & Raymond J. Gorte

Department of Chemical Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

The direct electrochemical oxidation of dry hydrocarbon fuels to generate electrical power has the potential to accelerate substantially the use of fuel cells in transportation and distributed-power applications¹. Most fuel-cell research has involved the use of hydrogen as the fuel, although the practical generation and storage of hydrogen remains an important technological hurdle². Methane has been successfully oxidized electrochemically^{3–6}, but the susceptibility to carbon formation from other hydrocarbons that may be present or poor power densities have prevented the application of this simple fuel in practical applications¹. Here we report the direct, electrochemical oxidation of various hydrocarbons (methane, ethane, 1-butene, *n*-butane and toluene) using a solid-oxide fuel cell at 973 and 1,073 K with a composite anode of copper and ceria (or samaria-doped ceria). We demonstrate that the final products of the oxidation are CO₂ and water, and that reasonable power densities can be achieved. The observation that a solid-oxide fuel cell can be operated on dry hydrocarbons, including liquid fuels, without reforming, suggests that this type of fuel cell could provide an alternative to hydrogen-based fuel-cell technologies.

The principle of operation of a solid-oxide fuel cell (SOFC) involves reduction of molecular O₂ at the cathode, diffusion of the O^{2–} through a zirconia-based electrolyte, and oxidation of the fuel by O^{2–} at the anode⁷. SOFCs must operate at high temperatures (to enable diffusion of O^{2–} through the electrolyte), so that C–H bond activation in hydrocarbons is feasible. The primary difficulty encountered during direct oxidation of hydrocarbons is rapid deactivation due to carbon deposition on the anode. The anode in conventional designs is a composite of Ni and yttria-stabilized zirconia (YSZ); this composite is an electronic conductor (due to Ni) and also an ionic conductor (due to YSZ). However, Ni catalyses formation of graphite from hydrocarbons: except for a narrow range of operating temperatures and only for methane³, carbon formation with Ni-based anodes is unavoidable.

We have recently shown^{5,8} that excellent performance can still be achieved if the Ni-YSZ anode is replaced by a Cu-ceria anode. Cu is

an excellent electronic conductor but a poor catalyst for C–H bond activation; therefore, carbon formation, which limits the use of Ni cermet, is avoided with the Cu cermet. (The ceramic and metal composite is known as a cermet.) Ceria is used in the anode because of its high activity for hydrocarbon oxidation and its high ionic conductivity⁹. The SOFCs used here were prepared according to refs 5 and 8, with a 60-μm-thick, YSZ electrolyte, 12.5 mm in diameter, and a cathode formed from a 50:50 mixture of YSZ and La_{0.8}Sr_{0.2}MnO₃ powders. The anodes were 40 wt% Cu and 20 wt% CeO₂, held in place by a YSZ matrix formed from zircon fibres. In a second cell, a 20% Sm₂O₃–80% CeO₂ mixed oxide replaced CeO₂. Electronic contacts were formed using a Pt mesh and Pt paste. (On the basis of our previous work, in which cells without either Cu or ceria showed very poor performance⁵, we do not believe the Pt played a catalytic role; however, this clearly needs to be checked.) Finally, the electrolyte disk was sealed onto alumina tubes. The total surface area of the cathode and anode was 0.25 cm². The cathode was left open to air, and fuel at a total pressure of 1.01 bar was allowed to flow over the anode. An on-line gas chromatograph (GC) was used to measure the composition of the effluent product leaving the anode side of the fuel cell.

Figure 1 shows the performance of the cell with the Cu-ceria anode when supplied with H₂ or *n*-butane at 973 or 1,073 K. While the open circuit voltages (OCVs) for H₂ were close to the value predicted by the Nernst equation, 1.05 V, the OCV for *n*-butane was ~0.9 V, which is lower than the potential of 1.1 V calculated for complete oxidation. As GC analysis of the effluent from the anode showed negligible amounts of CO and CO₂ at open circuit, leaks are an unlikely cause of the low OCV. As pointed out by Marina and Mogensen¹⁰, the direct, electrochemical oxidation of hydrocarbons is unlikely to take place in one step. Therefore, we suggest that the low OCV for butane results from equilibrium involving partial oxidation reactions. For example, the oxidation of ethane to ethene and water, one of many possible reactions, has a theoretical OCV of 0.93 V.

The maximum power densities were higher for H₂ than for *n*-butane, with values of 0.31 W cm^{–2} and 0.18 W cm^{–2}, respectively, for H₂ and *n*-butane at 1,073 K, and 0.22 W cm^{–2} and 0.12 W cm^{–2} at 973 K. Significantly higher power densities can certainly be achieved with thin-film electrolytes³.

The crucial test is whether carbon formation occurs on the Cu-cermet anode when *n*-butane is used. This cell was operated at the maximum power density of 0.12 W cm^{–2} in dry butane at 973 K for a period of 48 hours with no observable change in performance. The power output was also stable at 1,073 K, but gas-phase reactions formed tar on the walls of the alumina tube at the

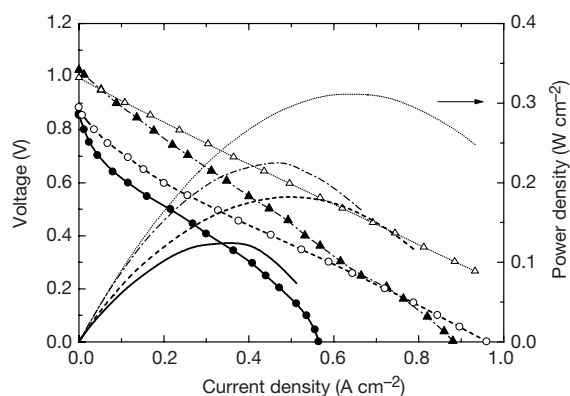


Figure 1 Power densities and current density–voltage relationships for an SOFC using the Cu-ceria composite anode. The cell had a 60-μm electrolyte, and data are shown for the following fuels: filled circles, *n*-butane at 973 K; open circles, *n*-butane at 1,073 K; filled triangles, H₂ at 973 K; and open triangles, H₂ at 1,073 K.

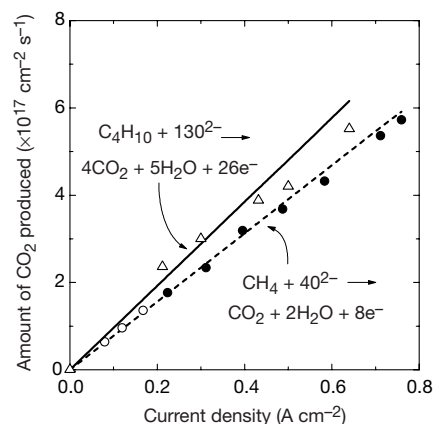


Figure 2 CO₂ production for methane and *n*-butane as a function of current density using the Cu-ceria composite anode. The data were obtained by gas analysis for: *n*-butane at 973 K (open triangles), methane at 973 K (open circles), and methane at 1,073 K (filled circles). The lines were calculated from the current density, assuming complete oxidation.

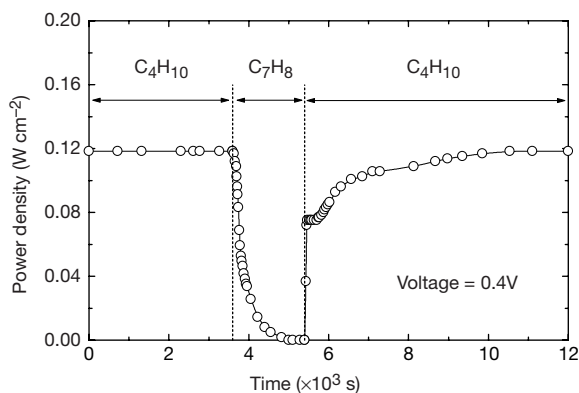
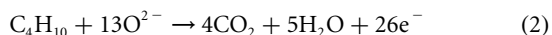


Figure 3 Effect of switching fuel type on the cell with the Cu-ceria composite anode at 973 K. The power density of the cell is shown as a function of time. The fuel was switched from *n*-butane (C₄H₁₀) to toluene (C₇H₈), and back to *n*-butane.

higher temperature. Visual inspection of a cell after two days in *n*-butane at 1,073 K showed that the anode itself remained free of the tar deposits that covered the alumina walls.

Although it is possible that the power generated from *n*-butane fuels resulted from oxidation of H₂—formed by gas-phase reactions of *n*-butane that produce hydrocarbons with a lower C:H ratio—other evidence shows that this is not the case. First, experiments were conducted in which the cell was charged with *n*-butane and then operated in a batch mode without flow. After 30 minutes of batch operation with the cell short-circuited, GC analysis showed that all of the *n*-butane in the cell had been converted completely to CO₂ and water. (Negligible amounts of CO₂ were formed in a similar experiment with an open circuit.) Second, analysis of the CO₂ formed under steady-state flow conditions, shown in Fig. 2, demonstrates that the rate of CO₂ formation increased linearly with the current density. (It was not possible for us to quantify the amount of water formed in our system.) Figure 2 includes data for both *n*-butane at 973 K, and methane at 973 K and 1,073 K. The lines in the figure were calculated assuming complete oxidation of methane (the dashed line) and *n*-butane (the solid line) to CO₂ and water according to reactions (1) and (2):



With methane, only trace levels of CO were observed along with CO₂, so that the agreement between the data points and the calculation demonstrates consistency in the measurements and no leaks in the cell. With *n*-butane, simultaneous, gas-phase, free-radical reactions to give hydrocarbons with various C:H ratios make quantification more difficult; however, the data still suggest that complete oxidation is the primary reaction. Furthermore, the batch experiments show that the secondary products formed by gas-phase reactions are ultimately oxidized as well. Taken together, these results demonstrate the direct, electrocatalytic oxidation of a higher hydrocarbon in a SOFC.

Along with our observation of stable power generation with *n*-butane for 48 hours, Fig. 3 further demonstrates the stability of the composite anodes against coke formation. Aromatic molecules, such as toluene, are expected to be precursors to the formation of graphitic coke deposits. In Fig. 3, the power density was measured at 973 K and 0.4 V while the fuel was switched from dry *n*-butane, to 0.033 bar of toluene in He for 30 minutes, and back to dry *n*-butane. The data show that the performance decreased rapidly in the presence of toluene. Upon switching back to dry *n*-butane, however,

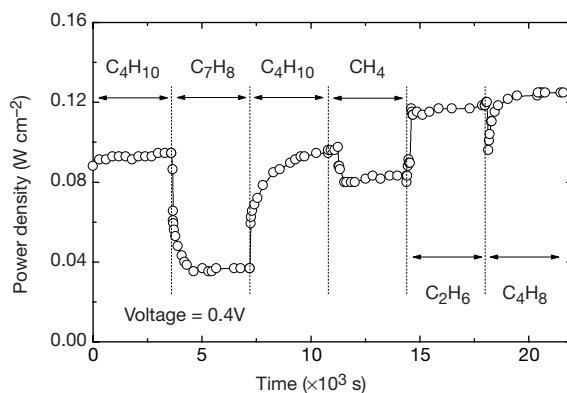


Figure 4 Effect of switching fuel type on the cell with the Cu-doped ceria composite anode at 973 K. The power density is shown as a function of time. The fuels were: *n*-butane (C₄H₁₀), toluene (C₇H₈), *n*-butane, methane (CH₄), ethane (C₂H₆), and 1-butene (C₄H₈).

the current density returned to 0.12 W cm⁻² after one hour. Because the return was not instantaneous, it appears that carbon formation occurred during exposure to toluene, but that the anode is self-cleaning. We note that the electrochemical oxidation of soot has been reported by others¹¹.

The data in Fig. 4 show that further improvements in cell performance can be achieved. For these experiments, samaria-doped ceria was substituted for ceria in the anode, and the current densities were measured at a potential of 0.4 V at 973 K. The power densities for H₂ and *n*-butane in this particular cell were approximately 20% lower than for the first cell, which is within the range of our ability to reproduce cells. However, the power densities achieved for some other fuels were significantly higher. In particular, stable power generation was now observed for toluene. Similarly, Fig. 4 shows that methane, ethane and 1-butene could be used as fuels to produce electrical energy. The data show transients for some of the fuels, which are at least partially due to switching.

The role of samaria in enhancing the results for toluene and some of the other hydrocarbons is uncertain. While samaria is used to enhance mixed (ionic and electronic) conductivity in ceria and could increase the active, three-phase boundary in the anode, samaria is also an active catalyst¹². Other improvements in the performance of SOFCs are possible. For example, the composite anodes could be easily attached to the cathode-supported, thin-film electrolytes that have been used by others to achieve very high power densities³. In addition to raising the power density, thinner electrolytes may also allow lower operating temperatures.

Additional research is clearly necessary for commercial development of fuel cells which generate electrical power directly from hydrocarbons; however, the work described here suggests that SOFCs have an intriguing future as portable, electric generators and possibly even as energy sources for transportation. The simplicity afforded by not having to reform the hydrocarbon fuels is a significant advantage of these cells. □

Received 13 September 1999; accepted 26 January 2000.

1. Steele, B. C. H. Running on natural gas. *Nature* **400**, 620–621 (1999).
2. Service, R. F. Bringing fuel cells down to earth. *Science* **285**, 682–685 (1999).
3. Perry Murray, E., Tsai, T. & Barnett, S. A. A direct-methane fuel cell with a ceria-based anode. *Nature* **400**, 649–651 (1999).
4. Putna, E. S., Stubenrauch, J., Vohs, J. M. & Gorte, R. J. Ceria-based anodes for the direct oxidation of methane in solid oxide fuel cells. *Langmuir* **11**, 4832–4837 (1995).
5. Park, S., Craciun, R., Vohs, J. M. & Gorte, R. J. Direct oxidation of hydrocarbons in a solid oxide fuel cell: I. methane oxidation. *J. Electrochem. Soc.* **146**, 3603–3605 (1999).
6. Steele, B. C. H., Kelly, I., Middleton, P. H. & Rudkin, R. Oxidation of methane in solid-state electrochemical reactors. *Solid State Ionics*, **28**, 1547–1552 (1988).
7. Lloyd, A. C. The power plant in your basement. *Sci. Am.* **281**(1), 80–86 (1999).

8. Craciun, R. *et al.* A novel method for preparing anode cermets for solid oxide fuel cells. *J. Electrochem. Soc.* **146**, 4019–4022 (1999).
9. Trovarelli, A. Catalytic properties of ceria and CeO₂-containing materials. *Catal. Rev. Sci. Eng.* **38**, 439–520 (1996).
10. Marina, O. A. & Mogensen, M. High-temperature conversion of methane on a composite gadolinia-doped ceria-gold electrode. *Appl. Catal. A* **189**, 117–126 (1999).
11. Christensen, H., Dinesen, J., Engel, H. H. & Hansen, K. K. *Electrochemical Reactor For Exhaust Gas Purification 1–5* (Society of Automotive Engineers paper no. 1999-01-0472, Warrendale, Philadelphia, 1999).
12. Yamanaka, I. & Otsuka, K. Catalysis of Sm³⁺ for the oxidation of alkanes with O₂ in the liquid phase. *J. Mol. Catal. A* **95**, 115–120 (1995).

Acknowledgements

We thank C. Wang and W. L. Worrell for technical advice. This work was supported by the Gas Research Institute.

Correspondence and requests for materials should be addressed to R.J.G. (e-mail: gorte@seas.upenn.edu).

Single crystals of an ionic anthracene aggregate with a triplet ground state

H. Bock*, K. Gharagozloo-Hubmann*, M. Sievert*, T. Prisner* & Z. Havlas†

* Department of Chemistry, Johann Wolfgang Goethe University, Marie Curie Strasse 11, D-60439 Frankfurt am Main, Germany

† Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Flemingovo Nam 2, CZ-16610 Prague, Czech Republic

Crystalline supramolecular aggregates consisting of charged organic molecules, held together through metal-cluster-mediated Coulomb interactions, have attracted interest owing to their unusual structural, chemical and electronic properties^{1–3}. Aggregates containing metal cation clusters ‘wrapped’ by lipophilic molecular anions have, for example, been shown^{4,5} to be kinetically stable and soluble in nonpolar liquids such as saturated hydrocarbons. The formation of supramolecular aggregates can even be exploited to generate aromatic hydrocarbons that carry four negative charges and crystallize in the form of organic poly(metal cation) clusters^{6,7} or helical polymers⁸. Here we report the anaerobic crystallization of an ionic organic aggregate—a contact ion septuple consisting of a fourfold negatively charged ‘tripledecker’ of three anthracene molecules bridged by four solvated potassium cations. Its electronic ground state is shown experimentally, using temperature-dependent electron paramagnetic resonance spectroscopy, to be a triplet. Although the spins in this biradical ionic solid are separated by a considerable distance, density functional theory calculations⁹ indicate that the triplet ground state is 84 kJ mol^{–1} more stable than the first excited singlet state. We expect that the successful crystallization of the ionic solid we report here, and that of a covalent organic compound with a triplet ground state¹⁰ at room temperature, will

stimulate further attempts to develop new triplet-ground-state materials for practical use.

Molecular anions M^{n–} and their counteranions Met⁺_{sol} (where Met is the alkali metal) crystallize from aprotic solvents either as solvent-separated [M^{n–}][Met⁺_{sol}]_n or as solvent-shared [M^{n–}Met⁺_n]_{sol} ion pairs^{11,12}. They are formed in a multidimensional network of solution equilibria¹³ involving electron transfer and contact ion pair formation as well as aggregation, and are often dominated by cation solvation¹⁴; such solvation is known to control numerous chemical reactions, including geochemical and biochemical reactions. For instance, if the π -hydrocarbon anthracene is reduced in ether or alkylamine solutions at alkali-metal surfaces [Met]_s, slight modifications of the conditions lead to various reduction products which can be structurally characterized in single crystals^{15,16} (Fig. 1).

The manifold of possible anthracene anion salt structures (Fig. 1) ranges from the solvent-separated sodium salt of the bare anthracene radical anion¹⁵ (Fig. 1a), via the double η^6 -half-sandwiches of two [Li⁺TMEDA] ligands on both faces of the anthracene dianion¹⁶ (Fig. 1b), to the ion quadruple of two anthracene radical anions connected by a [K⁺THF]₂⁺² bridge (Fig. 1c) (see Methods), and stimulated attempts to crystallize even larger aggregates. (THF, tetrahydrofuran; TMEDA, tetramethylethylenediamine.)

Low-gradient crystallization by diffusion of n-hexane from an added layer into the blue THF solution of anthracene that is reduced at a potassium metal mirror yields—after three days standing at room temperature—black needles with a violet lustre (see Methods). On further standing of the decanted solution, another batch of black needles is collected. X-ray structure analysis proves the first crystal fraction to be the title compound (that is, the contact ion septuple [(anthracene)₃^{4–}(K⁺THF)₂⁺²]₁ (Fig. 2), whereas the second batch contains the smaller aggregate [(anthracene)₂^{2–}(K⁺THF)₂⁺²]₁ (Fig. 1c). This suggests that the smaller compound might act as a ‘seed’, from which the crystal of the three-layered anthracene tetraanion (Fig. 2) might grow.

The crystal structure analysis of the anthracene potassium septuple (Fig. 2) reveals a hitherto unknown type of hydrocarbon aggregate: in the centrosymmetric double sandwich, the planar anthracene anions exhibit centroid/centroid distances of 565 pm and are tilted by interplanar angles of +54° and –54°. The four K⁺ centres (with a total coordination number of 14) are each η^6 -connected to the peripheral six-membered rings with contacts K⁺...C^{δ–} of 312–322 pm length to the outer anthracene ligands, and of 290–335 pm to the central anthracene; the perpendicular distances K⁺...ring amount to 289 and 275 pm. (C^{δ–} indicates partial charges at the individual centres.) The K⁺ centres are 431 pm apart, and each one exhibits two contacts K⁺...O to the O centres of two THF solvate ligands of lengths 265 and 285 pm. The planar anthracene anion rings are distorted differently: relative to the neutral molecule, the bonds C1–C2 are shortened by 4 pm and C2–C3 are elongated by 2 pm for the outer ligands, or shortened by 5 pm and elongated by 4 pm for the central dianionic bridge. This structural diversity is already known from published anthracene anion structures^{15,16} (Fig. 1), and suggests that the outer ligands are

Table 1 Density functional theory calculations

	Singlet	Triplet
ΔE_{total} (kJ mol ^{–1})		–84
Q (M ^{2–})	–1.58	–1.52
Q (M [–])	–0.86	–0.83
ρ (M [–])	0	0.97

Selected results from density functional theory (DFT) calculations for the triplet ground state and the lowest singlet state of the biradical salt [(anthracene)^{2–}(K⁺THF)₂]₁. Shown are difference in total energies ΔE_{total} and natural bond orbital charges Q as well as spin distribution ρ in the inner (M^{2–}) or the outer (M[–]) anthracene ligands.

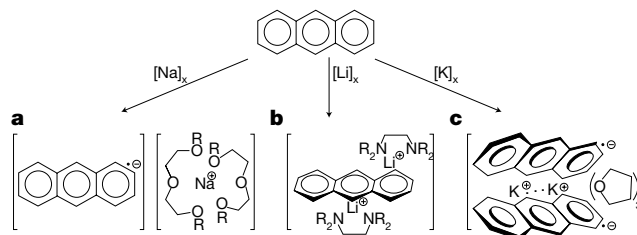


Figure 1 Selected examples of anthracene anion salts. **a**, Anthracene radical anion-sodium bis(triglyme) (ref. 15); **b**, anthracene dianion-dilithium-bis(TMEDA) (ref. 16); **c**, [(anthracene)₂^{2–}(K⁺THF)₃⁺²]₁ (see Methods).

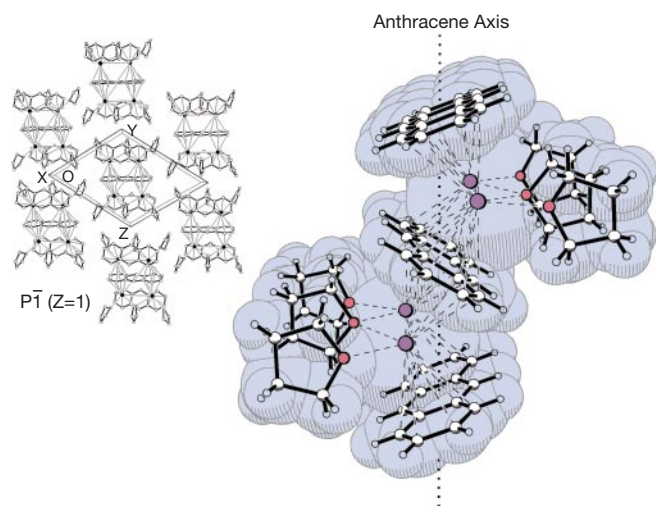


Figure 2 Single-crystal structure of $[(\text{anthracene})_3^-(\text{K}_2\text{THF}_3)^{+2} \text{anthracene}^{2-} (\text{K}_2\text{THF}_3)^{+2} (\text{anthracene}^-)]$ at 150 K. Shown is the unit cell $P\bar{1}$ ($Z = 1$) with adjacent aggregates, and its space-filled as well as skeletal structures (colour code: C and H, white; O, red; K violet). In the text, this compound is referred to as $[(\text{anthracene})_3^-(\text{K}_2\text{THF}_3)_2]_1$, for brevity; the longer formula is used here to aid comparison with the space-filling structure.

anthracene radical anions, whereas the central anthracene unit should rather be a dianion. The structure that we determined for the anthracene potassium quintuple (Fig. 1c), the potential 'seed' intermediate that was crystallized separately from the same solution (see Methods), shows many similarities: the interplanar tilting angle between the two anthracene radical anions amounts to 57° and their centroid/centroid distance to 490 pm.

Starting from the crystal structure coordinates, we performed unrestricted density functional theory (DFT) UB3LYP/6-31G* calculations for both the 154-atom contact ion septuple (Fig. 2) and the 89-atom contact ion quintuple (Fig. 1c) in both their singlet and their triplet states (see Methods). Unexpectedly, in both cases the triplet states were predicted to be more stable, by 84 kJ mol^{-1} for $[(\text{anthracene})_3^-(\text{K}_2\text{THF}_3)_2]_1$ (Fig. 2 and Table 1) and by 13 kJ mol^{-1} for $[(\text{anthracene})_2^-(\text{K}_2\text{THF}_3)^{+2}]_1$ (Fig. 1c).

The DFT/natural bond orbital charges that we calculated (Table 1) differ only slightly for the singlet and the triplet state (Fig. 1), but the spin is predicted to be located almost exclusively in the two outer anthracene radical anion ligands, the centroids of which are 1,030 pm apart, on both sides of the central anthracene dianion mediating the spin-spin interaction.

Different spin states S can be distinguished with pulsed electron paramagnetic resonance (EPR) measurement by their distinct Rabi oscillation frequency under resonant coherent microwave excitation¹⁷. Theoretically, the Rabi frequency scales with $\sqrt{S(S+1) - m_s(m_s+1)}$ for a transition from m_s to m_{s+1} .

We compared the observed Rabi frequency for the anthracene crystals with standard calibration samples (see Fig. 3 and Methods), and demonstrated clearly that the observed signal from the crystals corresponds to a triplet state ($S = 1$). For such a triplet spin state, a typical dipolar splitting of the EPR line by $(D/r^3)(3\cos^2\theta - 1)$ (where D is the dipolar coupling constant, r is the spin-spin distance, and θ is the angle between dipolar vector and external magnetic field) would be expected to be directly observable in the spectrum. Instead, only a narrow homogeneous line (line-width $\Delta\omega_{1/2} = 2/T_2 = 2/T_1$ with T_1 longitudinal and T_2 transversal relaxation times) without any dipolar and hyperfine contribution was observed. The reason that the dipolar coupling of the triplet spin

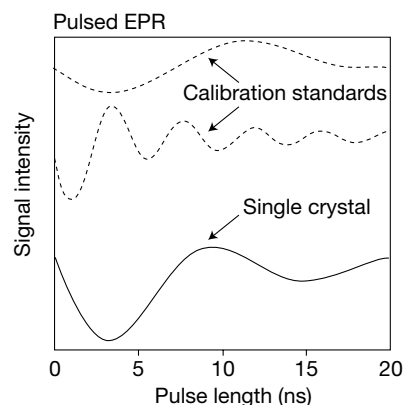


Figure 3 Rabi oscillations of the EPR signal of a single crystal of $[(\text{anthracene})_3^-(\text{K}_2\text{THF}_3)_2]_1$. The oscillation frequencies have been calibrated by two standard samples; one with a spin-state $S = 1/2$, the other with a spin-state $S = 5/2$ (see Methods for details). All measurements were performed under identical conditions (same Q value of microwave cavity, microwave excitation power, sample size and temperature). The observed Rabi frequencies were 7 MHz for the $S = 1/2$ standard, 9.5 MHz for the anthracene crystals, and 24 MHz for the $S = 5/2$ standard. These measurements show clearly that the spin state of the observed signal corresponds to a triplet state (see Methods for rationale).

cannot be observed in the continuous-wave EPR spectra might be due to a strong exchange coupling of the triplet states in the crystal, which would narrow the EPR signal to a homogeneous line.

The crystallization of the three-layered trianthracene tetraanion potassium salt, as well as its predicted and proven triplet ground state at room temperature, are of general importance both for multielectron reduction processes of molecules in solution and for the design of new biradical ionic solids. We conclude that the sterically limited solvation by THF, known for Na^+ counteranions¹⁴, favours the self-aggregation within the solvent-dependent equilibria network of electron transfer and contact ion formation¹³. From this solution both the anthracene trimer (Fig. 2) as well as the dimer biradical (Fig. 1c) crystallize selectively. The trianthracene contact ion septuple described here seems to be the first crystalline organic ionic material for which a triplet biradical ground state has been calculated and confirmed by EPR measurement. The structure of a neutral triplet biradical, tris(3,5-di(*t*-butyl)-4-oxophenylene) methane at room temperature, has been reported before¹⁰.

We consider that the crystallization and structural characterization of covalent as well as ionic organic compounds with a triplet ground state should stimulate further investigations. In particular, our work suggests that the synthesis of air-stable biradicals with lipophilically shielding paramagnetic organometallic ligands on both ends of a suitable spin propagating transducer seems to be feasible. Also, referring especially to the exchange coupling between adjacent aggregates within the crystal lattice, the exploration of the physical properties as well as the potential technical applications of new triplet-ground-state materials could be rewarding. □

Methods

Preparation and crystal growth of compounds

From 140 mg (7.6 mmol) potassium, a metal mirror is generated on the wall of a dried Schlenk trap by gentle heating at 10^{-5} mbar. Addition of 310 mg (1.7 mmol) of anthracene in 10 ml aprotic tetrahydrofuran (THF) ($c_{\text{H}_2} < 1 \text{ p.p.m.}$) produces a blue solution, which after 1 day of standing at room temperature is covered with a layer of 24 ml *n*-hexane. After another 3 days, black crystals with a violet lustre of $[(\text{anthracene})_3^-(\text{K}_2\text{THF}_3)_2]_1$ (Fig. 2) are collected from the trap wall near the phase border line. Keeping the solution for another 7 days at room temperature yields black crystals with a violet lustre of $[(\text{anthracene})_2^-(\text{K}_2\text{THF}_3)^{+2}]_1$ (Fig. 1c).

X-ray structure analyses

$[(\text{anthracene})_3^+(\text{K}_2^+\text{THF}_3)_2^-]_1$ (Fig. 2). Space group $P\bar{1}$, $a = 9.388(2)$, $b = 13.265(2)$, $c = 14.192(2)$ Å, $\alpha = 62.87(6)^\circ$, $\beta = 85.20(1)^\circ$, $\gamma = 70.47(1)^\circ$, $V = 1477.5(5)$ Å³, $Z = 1$, $\rho = 1.263$ g cm⁻³, $\mu = 0.35$ mm⁻¹, $F(000) = 598$, crystal size = $0.4 \times 0.35 \times 0.35$ mm; crystals were removed from the trap wall after 4 days under argon, immediately immersed in perfluoropolyether oil and mounted under a flow of nitrogen cooled to 150 K on a Siemens P4 diffractometer (Mo K α , graphite monochromator). A total of 6,211 reflections ($2.31 < \theta < 26.27^\circ$) were collected, of which 4,911 unique reflections ($R_{\text{int}} = 0.0308$) were used. The structure was solved using the program SHELXS-93 and refined (343 parameters) using the program SHELXS-93 to $R = 0.0434$ for 3,941 reflections with $I > 2\sigma(I)$ (both programs from G. M. Sheldrick, Univ. Göttingen, 1993).

$[(\text{anthracene})_2^+(\text{K}_2^+\text{THF}_3)_2^-]_1$ (Fig. 1c). Space group $C2/c$, $a = 14.186(3)$, $b = 10.633(2)$, $c = 23.382(4)$ Å, $\beta = 103.78(1)^\circ$, $V = 3425.4(11)$ Å³, $Z = 4$, $\rho = 1.262$ g cm⁻³, $\mu = 0.314$ mm⁻¹, $F(000) = 1,384$, crystal size = $0.25 \times 0.62 \times 0.84$ mm; crystals were removed from the trap wall after 11 days under argon, immediately immersed in perfluoropolyether oil and mounted under a flow of nitrogen cooled to 150 K on a Siemens P4 diffractometer (Mo K α , graphite monochromator). A total of 3,783 reflections ($2.42 < \theta < 25^\circ$) were collected, of which 2,995 unique reflections ($R_{\text{int}} = 0.0337$) were used. The structure was solved using the program SHELXS-93 and refined (203 parameters) using the program SHELXS-93 to $R = 0.0507$ for 1,801 reflections with $I > 2\sigma(I)$ (both programs from G. M. Sheldrick, Univ. Göttingen, 1993). The β -carbon centres of the THF ligand in the middle position are disordered over two positions with site occupation factors of 0.49 and 0.51.

Unrestricted density functional theory calculations

For both the rather large aggregates $[(\text{anthracene})_3^+(\text{K}_2^+\text{THF}_3)_2^-]$ and $[(\text{anthracene})_2^+(\text{K}_2^+\text{THF}_3)_2^-]$, the singlet- and triplet-state computations have been performed at the UB3LYP/6-31G* level, using the NEC SX4 supercomputer of the Höchstleistungs-Rechenzentrum Stuttgart, Germany using the Gaussian 94 program package⁹. The singlet-state calculations were started from the broken symmetry orbital. The total energies of the triplet states are calculated to be -5384.72797 a.u. and -2976.24778 a.u. The charge and spin distributions were obtained by natural bond orbital analysis after enlarging the capacity of the corresponding program module.

EPR measurements

Pulsed and continuous-wave EPR measurements were performed with a Bruker E580 ELEXIS X-band spectrometer. Single crystals ($4 \times 0.35 \times 0.35$ mm) were sealed in 3-mm-diameter X-band quartz EPR capillaries under argon in n-hexane solution. The single crystals give rise to a narrow homogeneous EPR line, with a linewidth of 1.9 G and a g value close to 2, typical for a π -delocalized organic radical. No dipolar splitting of the signal could be observed in the temperature range 10–290 K. The Rabi oscillation behaviour of the signal¹⁷ of these crystals have been compared with a spin-state $S = 1/2$ radical salt of perylene ((perylene)₂(AsF₆)_{0.75}(PF₆)_{0.35}·0.85 CH₂Cl₂) and a spin-state $S = 5/2$ MnO sample (only the central $m_S = -1/2$ to $+1/2$ transition observed) under identical conditions (same Q value of microwave cavity, microwave excitation power, sample size and temperature). The theoretical predicted ratio of the oscillation frequencies are $1 : \sqrt{2} : 3$ for a $S = 1/2$, a $S = 1$ and the $m_S = -1/2$ to $+1/2$ transition of a $S = 5/2$ spin state. The observed Rabi frequencies were 7 MHz for $S = 1/2$, 9.5 MHz for the anthracene crystals, and 24 MHz for the MnO sample. Therefore these measurements clearly showed that the spin state of the observed signal corresponds to a triplet state. The temperature dependence (10–290 K) of the EPR signal as well as the absolute signal intensity exclude an assignment of the signal to a low-lying excited triplet state or to impurities or defects. Taken together, these measurements prove that the ground state of this compound is indeed a triplet spin state.

Received 20 July 1999; accepted 17 January 2000.

- Lehn, J. M. *Supramolecular Chemistry, Concepts and Perspectives* (VCH, Weinheim, 1995).
- Müller, A., Reuter, H. & Dillingers, S. Supramolecular inorganic chemistry: small guests in small and large hosts. *Angew. Chem. Int. Edn Engl.* **34**, 2326–2361 (1995).
- Müller, A., Shah, S. Q. N., Böge, H. & Schmidtman, M. Molecular growth from a Mo₁₇₆ to a Mo₂₄₈ cluster. *Nature* **397**, 48–50 (1999).
- Bock, H., Lehn, J. M., Pauls, J., Holl, S. & Krenzel, V. Sodium salts of the bipyridine dianion polymer $[(\text{bpy})_2^2-[\text{Na}^+(\text{dme})_2]_x \text{ cluster } [(\text{Na}_2\text{O})_6^{4-} \text{Na}_4^+(\text{tmeda})_6^-]$ and monomer $[(\text{bpy})_2^2-[\text{Na}^+(\text{pmdta})_2]_2]$. *Angew. Chem. Int. Edn Engl.* **38**, 952–955 (1999).
- Bock, H. et al. The lipophilically wrapped polyanion aggregate $\{[\text{Ba}_6\text{Li}_3\text{O}_2]^{11+}[\text{OC}(\text{CH}_3)_3]_{11}(\text{OC}_6\text{H}_5)_6\}$, a face-sharing (octahedron + prismane) Ba₆Li₃O₂ polyhedron in a hydrocarbon ellipsoid: preparation, single crystal structure analysis, and density functional calculations. *Angew. Chem. Int. Edn Engl.* **34**, 1353–1355 (1995).
- Bock, H., Gharagozloo-Hubmann, K., Näther, C., Nagel, N. & Havlas, Z. $[(\text{Na}^+(\text{thf})_2)_4(\text{rubrene})^4-]$ crystallisation and structure determination of a contact ion quintuple for the first π -hydrocarbon-tetraanion. *Angew. Chem. Int. Edn Engl.* **35**, 631–633 (1996).
- Sekiguchi, A., Matsuo, T. & Kabuto, C. Synthesis and characterisation of the tetralithium salt of an octasilyl-substituted trimethylcyclopentadien-tetraanion. *Angew. Chem. Int. Edn Engl.* **36**, 2462–2464 (1997).
- Bock, H., Gharagozloo-Hubmann, K., Sievert, M. & Havlas, Z. Synthesis, crystal growth and structure determination of $[(6,13'\text{-Dipentacene}^{4-})(\text{K}^+\text{DME})_4]_n$: a twin-strand polymer helix. *Angew. Chem. Int. Edn Engl.* (submitted).
- Bock, H., Z. & Bock, H. Bare molecular anions of the saturated hydrocarbons: density functional charge and spin distribution based on their single crystal structures. *Collect. Czech. Chem. Commun.* **63**, 1245–1263 (1998).
- Bock, H., John, A., Havlas, Z. & Bats, J. W. The triplet biradical tris(3,5-di(tert-butyl)-4-oxophenyl-

lene)methane: crystal structure, spin and charge distribution. *Angew. Chem. Int. Edn Engl.* **32**, 416–418 (1993).

- Bock, H., Näther, C., Ruppert, K. & Havlas, Z. Tetraphenylbutadiene disodium dimethoxyethene: solvent-shared and solvent-separated ion triples within a single crystal. *J. Am. Chem. Soc.* **114**, 6907–6908 (1992).
- Bock, H., Näther, C., Ruppert, K. & Havlas, Z. Competing Na⁺ solvation: ether-shared and ether-separated triple ions of perylene dianion. *J. Am. Chem. Soc.* **117**, 3869–3870 (1995).
- Bock, H. et al. Distorted molecules: perturbation design, preparation and structures. *Angew. Chem. Int. Edn Engl.* **31**, 550–581 (1992).
- Bock, H., Näther, C., Havlas, Z., John, A. & Arad, C. Ether-solvated sodium ions in salts containing π -hydrocarbon anions: crystallisation, structures and semiempirical solvation energies. *Angew. Chem. Int. Edn Engl.* **33**, 875–878 (1994).
- Bock, H., Arad, C., Näther, C. & Havlas, Z. The structure of solvent-separated naphthalene and anthracene radical anions. *J. Chem. Soc. Chem. Commun.* 2393–2394 (1995).
- Rhine, W. E., Davis, J. & Stucky, G. Anthracene dianion dilithium bis(TMEDA). *J. Am. Chem. Soc.* **97**, 2079–2085 (1975).
- Stoll, S., Jeschke, G., Willer, M. & Schweiger, A. Nutation-frequency correlated EPR spectroscopy: the PEANUT experiment. *J. Magn. Reson.* **130**, 86–96 (1998).

Acknowledgements

We thank T. Hauck for the first crystal growth, C. Näther and S. Holl for structure determination, V. Krenzel for graphics, and J. T. Törling & H. Käß for EPR measurements. This work was supported by the Deutsche Forschungsgemeinschaft, the Fonds der Chemischen Industrie and the State of Hesse.

Correspondence and requests for materials should be addressed to H.B.

(e-mail: bock@www.anorg.chemie.uni-frankfurt.de). Further details of crystal growth and structure determination may be obtained from the Cambridge Crystallographic Data Centre (12 Union Road, Cambridge CB2 1EZ, UK) on quoting supplementary publication no. CCDC-137565 for the compound shown in Fig. 2, and no. CCDC-137566 for the compound shown in Fig. 1c.

Seismic hazard in the Marmara Sea region following the 17 August 1999 Izmit earthquake

Aurélia Hubert-Ferrari*, Aykut Barka†, Eric Jacques‡, Süleyman S. Nalbant§, Bertrand Meyer||, Rolando Armijo||, Paul Tapponnier|| & Geoffrey C. P. King||

* Department of Geosciences, Princeton University, Princeton, New Jersey 08544, USA

† ITU, Maden Fakültesi, Jeoloji Bölümü, Ayazaga, Istanbul, Turkey

‡ Observatoire des Sciences de la Terre de Strasbourg (UMS 830 of CNRS), Strasbourg, France

§ Geophysics Department, Engineering Faculty, Istanbul University, Istanbul, Turkey

|| Laboratoire de Tectonique et Mécanique de la Lithosphère (UMR 7578 of CNRS), Institut de Physique du Globe, Paris

On 17 August 1999, a destructive magnitude 7.4 earthquake occurred 100 km east of Istanbul, near the city of Izmit, on the North Anatolian fault. This 1,600-km-long plate boundary^{1,2} slips at an average rate of 2–3 cm yr⁻¹ (refs 3–5), and historically has been the site of many devastating earthquakes^{6,7}. This century alone it has ruptured over 900 km of its length⁶. Models of earthquake-induced stress change⁸ combined with active fault maps⁹ had been used to forecast that the epicentral area of the 1999 Izmit event was indeed a likely location for the occurrence of a large earthquake^{9,10}. Here we show that the 1999 event itself significantly modifies the stress distribution resulting from previous fault interactions^{9,10}. Our new stress models take into account all events in the region with magnitudes greater than 6 having occurred since 1700 (ref. 7) as well as secular interseismic stress change, constrained by GPS data¹¹. These models provide a consistent picture of the long term spatio-temporal behaviour of the North Anatolian fault and indicate that two events of magni-



tude equal to, or greater than, the Izmit earthquake are likely to occur within the next decades beneath the Marmara Sea, south of Istanbul.

Seventy kilometres east of Izmit, the North Anatolian fault (NAF) splits into two main segmented strands¹². The northern most active strand (NNAF) passes beneath the Sea of Marmara, while the

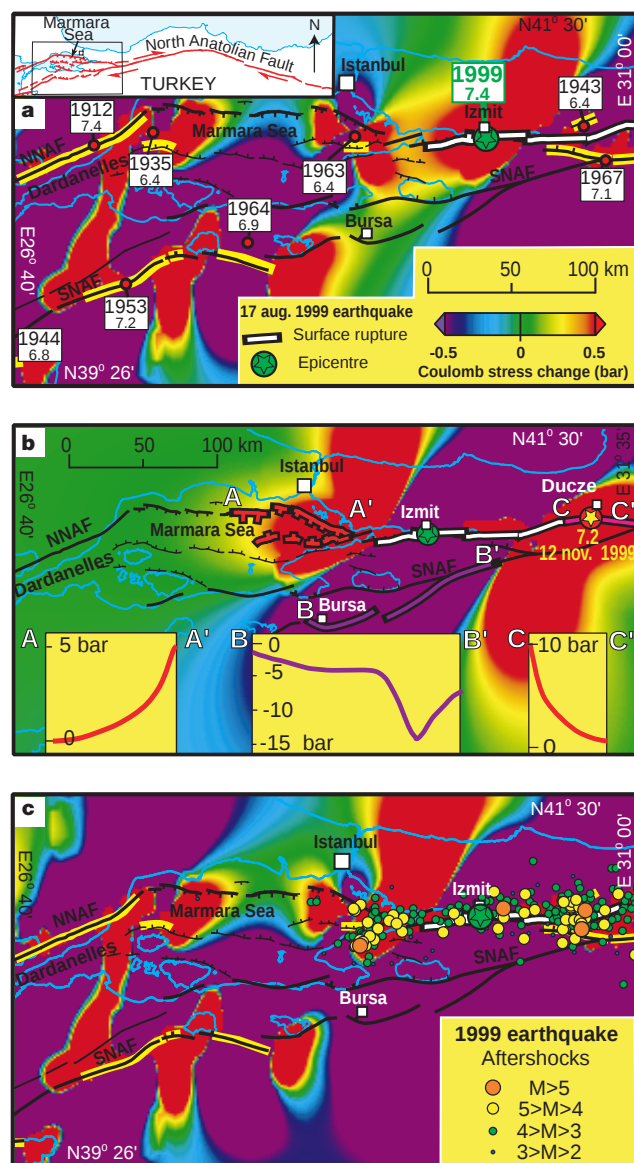


Figure 1 Stress changes due to earthquakes this century in the Marmara Sea area. These stress changes are shown at a depth of 6 km, and are calculated on optimally oriented faults (see Methods). **a**, Results for all eight $M > 6$ events before the 1999 Izmit earthquake⁹. Locations of fault elements used to model earthquakes are shown in yellow, and their epicentral location by a red dot. The area studied is shown in the inset. **b**, Stress induced by the 1999 Izmit earthquake that ruptured four on-land segments¹ (white). Rupture might have extended westward beneath the Gulf of Izmit, but is not modelled here. At the eastern tip of the Izmit rupture, fault segments (pink) between C and C' are strongly loaded by the Izmit event and ruptured during the $M = 7.2$ Duzce event on 12 November 1999. Two faults (red), west of the new rupture that bound the trough 20 km south of Istanbul are also strongly loaded. Insets show changes of Coulomb stress resolved on faults between A and A', B and B', and C and C', shown respectively in red, purple and pink. The single curve in the plot A–A' corresponds to both faults (red) as they happen to have almost identical stress loading. **c**, Stress changes due to all events since 1900. The aftershocks correlate with the stress changes, but with events notably absent in regions where the 1963 earthquake released stress.

southern strand (SNAF) remains on land south of the Sea^{5,9}. Since 1900 and before the Izmit event, five earthquakes of magnitude $6 \leq M < 7$ and three $M > 7$ earthquakes have ruptured fault segments of one or other of the strands⁹ (Fig. 1a). Each of the $M > 7$ earthquakes, whose fault parameters are constrained by instrumental data and surface rupture mapping, broke several segments⁹. They provide a guide to the assessment of slip during earlier $M > 7$ historical events.

The Coulomb stress^{8,13} field resulting from all earthquakes this century in the Marmara Sea region before the Izmit event is shown in Fig. 1a (ref. 9). The 1963 and 1967 earthquakes enhanced stress by 0.5–2 bar in the Izmit epicentral region^{9,10}. Over the last few years microseismicity had clustered around the August 1999 epicentre (Izinet network, Kandilli Observatory). The 1999 rupture nucleated in the area of increased stress, and extended along at least 110 km on four main segments (Fig. 1a) with dextral displacement of up to 5 m (ref. 1). As a result, the Coulomb stress resolved onto faults west of the rupture rose by 1–5 bar over a distance of 25 km, increasing the probability of an earthquake closer to Istanbul (Fig. 1b). At the eastern tip of the rupture, stress was increased by 10–0.6 bar along the Duzce fault, which then ruptured on 12 November 1999 with up to 5 m of dextral slip and locally up to 4 m of vertical offset. The events have also reduced stress by 3–15 bar along 120 km of the SNAF making an event there in the near future less likely (Fig. 1b). The largest aftershocks (Kandilli Observatory, preliminary locations) fall near the rupture or at rupture segment extremities where the Coulomb stress was enhanced (Fig. 1c). When earlier seismicity is included, most of the aftershocks at the western end of the rupture are confined to the area of Coulomb stress increase (Fig. 1c). This indicates that the small stress shadow of the $M = 6.4$ 1963 earthquake still persists after more than 36 years (refs 9, 13, 14).

The seismic behaviour of the NAF can be seen better in Coulomb models that include all $M > 6$ events since 1700 (see below). For the period 1700–1900, the information about damage is available. When combined with detailed mapping of Holocene faulting⁹ and scaling relations^{9,15} the modelling parameters (fault locations, dimensions, strike, dip, amplitude and rake of the slip vector) can be evaluated (see Methods). We also included tectonic loading in the modelling, a significant effect as the NAF has a high slip rate. We assume that associated localized shear occurs aseismically below a prescribed locking depth¹⁶, which we model using vertical dislocations positioned beneath the mean location of the surface faults and with slip vectors that fit local GPS data around the Marmara Sea (See Methods and Fig. 2). Both this tectonic loading and the

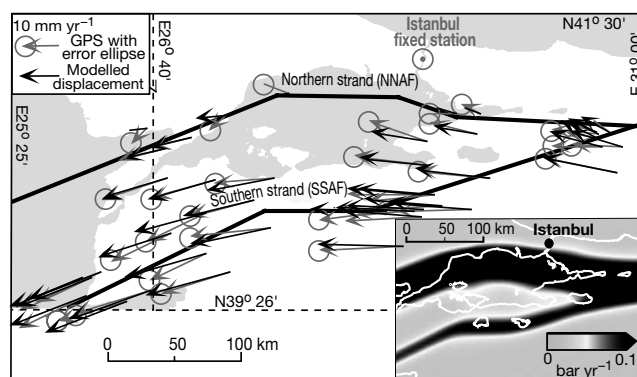


Figure 2 Model of the secular interseismic stress change. Black segments represent the location of the dislocation elements at depth, grey arrows show the observed GPS velocity field relative to the Istanbul station¹¹ (with circles indicating approximate daily repeatability errors¹¹) and black arrows show our model fit. The inset shows the form of the stress loading at depth of 6 km. Maximum loading is 0.4 bar yr^{-1} for the NNAF, and 0.1 bar yr^{-1} for the SNAF. NNAF, northern North Anatolian fault; SNAF, southern North Anatolian fault.

sequential loading of all $M > 6$ earthquakes since 1700 (excluding the Izmit event; Fig. 3a) are taken into account to calculate the stress distribution (Fig. 3b). Models do not take into account any viscous loading¹⁷. The stressed region where the Izmit earthquake occurred is restricted to 40 km, and includes the epicentre and all the western segments. The segments to the east were not uniformly stressed before the earthquake (Fig. 3b), but became stressed as failure proceeded. When the effect of the Izmit event is included (Fig. 3c), regions of enhanced stress are mainly restricted to the 160-km-long Marmara Sea trough. The faults of this complex pull-apart basin have now accumulated stress over 233 years (~ 100 bar) without any major earthquake, and thus should be approaching failure. Much of the character of the Coulomb stress distributions is dominated by large-scale slip deficits.

To explore further these large-scale features, we examine directly the accumulating slip release due to earthquakes on the NNAF (Fig. 3d). For the period 1700–1900, the component of co-seismic slip resolved parallel to relative plate motion³ is dominated by $M > 7$ events occurring between 1719 and 1766 (Fig. 3d). There are uncertainties in the magnitude of slip of about 1 m and the locations of segments which ruptured in each earthquake cannot be determined unambiguously, especially for the 1894 event (see discussion in Supplementary Information). This is also particularly true for the Marmara Sea pull-apart, where the fault geometry is poorly constrained: earthquakes could have ruptured north-dipping, south-dipping oblique faults or even pure strike-slip faults¹² that have a slightly different strike. However, all active faults are located in a zone 20–25 km wide, and Coulomb stress distributions for the different fault geometries at a distance from the fault plane remain similar⁹. In all the likely scenarios the westward propagating 1719–1766 sequence, which followed events farther to the east that ruptured about 500 km of the NAF in 1668¹⁸, released all the

stress previously accumulated. After 1766, stress began accumulating anew in the Marmara Sea region. After 146 years, 3.5 m of loading have accumulated on the NNAF. A single event then occurred in the west in 1912. In 1939, a major westward propagating sequence of earthquakes started rupturing the NAF from eastern Turkey to the eastern part of the Marmara Sea region ($M > 7$ earthquakes in 1957 and 1967, Fig. 3d)^{6,10}. When this sequence started, at least 5 m of loading had accumulated on the NAF since 1766 as a result of plate motion. The 17 August 1999 Izmit event fills part of a 5.5-m slip deficit on the NNAF (Fig. 3d). The 12 November 1999 Duzce event ($M = 7.2$) extends the 1999 Izmit rupture zone eastwards toward the NAF fault segment that ruptured in 1944. With this last event, all the active NAF segments have now ruptured from eastern Turkey to the Sea of Marmara.

There are three possible reasons why rupture did not continue towards the west. The first could be a shadow effect resulting from the 1963 earthquake (Fig. 1a), perhaps sufficient to arrest rupture. The second could be associated with the geometry of faulting. The predominantly strike-slip faults that ruptured in 1719 and 1999 do not extend far into the eastern Marmara Sea. Motion is transferred to fault segments with substantial components of normal slip that bound the trough south of Istanbul. Thus rupture may not readily propagate directly from one fault system to the other. Finally, the Coulomb model based on our understanding of segment geometry and slip history suggests that the rupture may have been arrested by a Coulomb minimum related to earlier events as well as to the 1963 event (Fig. 3b). When the Izmit event is included, a slip gap remains within the Marmara Sea over a distance of about 160 km where 5.5 m of slip has accumulated since 1766. Unlike the NNAF, the history along the SNAF since 1700 does not provide us with a period when the whole boundary has ruptured. However, major events have occurred also in sequence west of Bursa, and none have

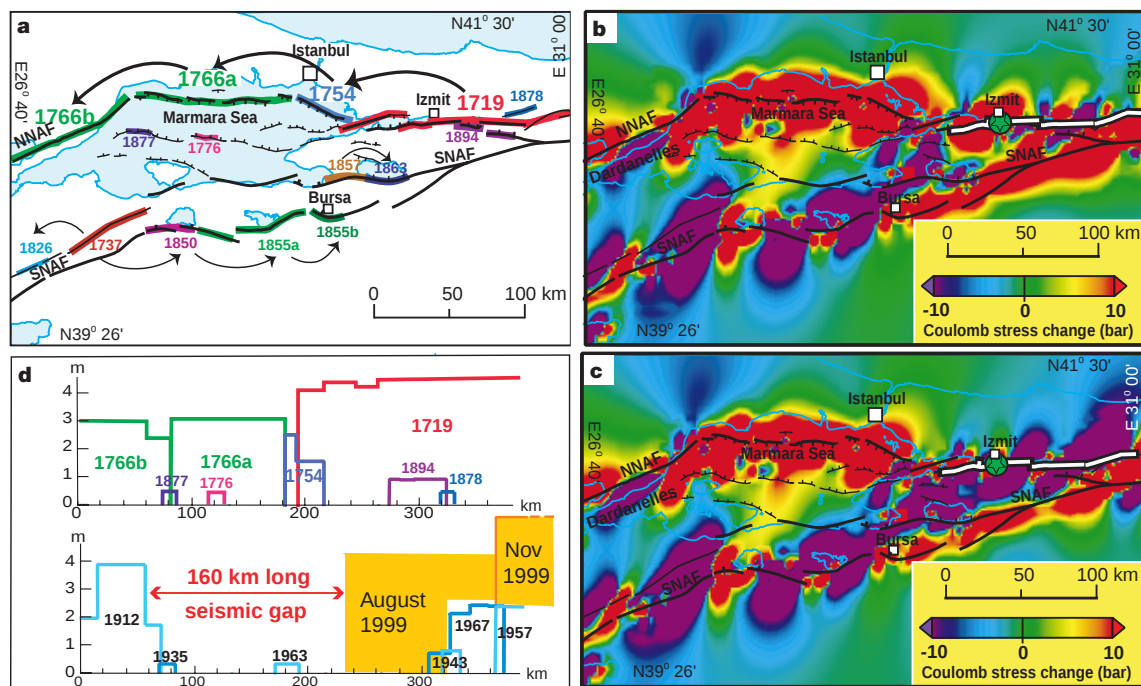


Figure 3 Earthquake-induced stress changes including historical seismicity since 1700. **a**, Location of fault elements used to model earthquakes between 1700 and 1900. **b**, Coulomb stress with tectonic loading distribution immediately before the Izmit earthquake. The 1999 earthquake epicentre is located in a high-stress area. **c**, Coulomb stress distribution with tectonic loading including the Izmit earthquake. The stresses are enhanced in the Marmara Sea area and reduced along the SNAF. **d**, Seismic slip resolved

parallel to relative plate motion³ for events on the NNAF and NAF since 1700 between longitude 26° 40' E and 31° 10' E. In the eighteenth century, the whole Marmara region slipped (top diagram) and by 1766 most of the accumulated slip would have been released. The smaller events contribute little to the total slip. Since 1900, major events have released slip in the east and west of the Marmara Sea region (bottom diagram). The slip due to the Izmit and Duzce events fills a gap to the east.

occurred to the east in this time period. The resulting stress accumulation can be seen in the Coulomb stress distributions (Fig. 3c). The last event to have ruptured the eastern part of the fault occurred in 1419 (ref. 7). Since then, 3.5 m of slip has accumulated over 125 km of the SNAF.

The above models depend on the accuracy of our data and on assumptions about the behaviour of faults in the Marmara Sea region. It might be supposed that our slip rate of 3 cm yr^{-1} is too high. However, the earthquake slip that has occurred in the Dardanelles region (3–4 m) and in the west of the Gulf of Izmit (4–5 m) is difficult to explain if the loading rate is less. It is possible that we have over-estimated both the slip in these historical earthquakes and the loading slip. A slip rate of 2 cm yr^{-1} on the NAF would reduce all slips and slip rates by one-third, but would preserve the spatial pattern of estimated historical slip (Fig. 3d). Similar proportional changes apply to stress amplitudes in the Coulomb models. Thus changing the assumed loading rate does not alter our models. The only way to significantly change the picture would be to assume high creep rates for faults beneath the Marmara Sea. The fact that these faults have hosted numerous earthquakes throughout history makes this improbable.

The significance of the observations here can be clarified by discussing the slip deficits and Coulomb stress changes together. A major slip gap (5.5 m since 1766) exists on faults beneath the Marmara Sea, and might be greater than the slip released by events in 1754 and 1766 following the 1719 earthquake (Fig. 3d). Faults in the west and east of the Marmara Sea region have experienced Coulomb stress increases due to events this century (Fig. 1). The stress increase in the west has been present since the 1912 earthquake while that in the east has just happened. In 1719 when an event similar to the Izmit event occurred, it was followed 35 years later (in 1754) by an event on a fault to the west. The same fault has now experienced a Coulomb stress increase of 1–5 bar (Fig. 1b). It is therefore likely that this fault will slip again and furthermore, the risk during the aftershock sequence must be regarded as particularly high^{10,19}. Following the 1754 earthquake, the 1766 events ruptured the rest of the fault 12 years later. It is very likely that a similar sequence will recur.

From the present study, the NAF earthquake behaviour could be summarized in the following way. Stress accumulates on the fault continuously as a result of plate motion. When the whole fault system is near failure, a critical stage is reached that is characterized by an extreme susceptibility to small perturbations and strong correlation between different parts of the system^{20,21}. Small stress increases (1–5 bar) are then sufficient to trigger failure in the upper crust, and the fault tends to rupture over most of its length in a cascading sequence of earthquakes. Between 1668 and 1766, the NAF seems to have ruptured over most of its length in such a sequence in about 100 years. This century the fault ruptured from eastern Turkey to the Izmit bay in a sequence of eight $M > 7$ events, the Izmit and the Duzce earthquakes being the last ones. In the next decades, it is likely that rupture of the fault segments of the Marmara Sea region will complete the sequence. □

Methods

Coulomb modelling

The stress fields due to earthquakes are calculated using dislocation theory^{8,22}. Dislocations to represent each event or sub-event are approximated by rectangular planes whose geometry, dimension and location must be known. Places of likely future failure are then identified as having an increased Coulomb stress ($\Delta\sigma_f$), which is calculated at a depth of 6 km (ref. 9; $\Delta\sigma_f = \Delta\tau - \mu' \Delta\sigma_n$ with $\mu' = 0.4$, where $\Delta\tau$ and $\Delta\sigma_n$ are respectively the change in shear and normal stresses on likely future fault planes). To specify preferred orientations of faulting, a regional stress is applied and the program searches for the optimal strike, dip and rake of the fault unless it is specified (horizontal principal compressive stress axis of $100\text{--}150$ bar oriented $N120^\circ E$)^{8,9}. All fault dislocations extend from the surface to a depth of 15 km, which is slightly greater than the 12.5 km previously used⁹ to be consistent with the loading model. In the program, $\Delta\tau = \Delta\epsilon_{13} + \Delta\epsilon_{31}$ (where $\Delta\tau$ is the shear stress on the target fault plane, and $\Delta\epsilon_{13}$ is the tensor component of shear stress rotated into that plane),

so the calculated Coulomb stresses are to a first approximation doubled when compared to previous work^{8,9,10} where $\Delta\tau = \Delta\epsilon_{13}$.

Tectonic loading

This is modelled by dislocations assumed to extend through the lithosphere to infinite depth (this is similar to a dislocation extending to 100 km and a fluid asthenosphere beneath). The slip parameters for these deep dislocations were established to fit the following criteria. The overall rate should be consistent with recent geological rates (2.0 cm yr^{-1})^{4,5}, geodetic rates ($2.5\text{--}3.0 \text{ cm yr}^{-1}$)³, a reasonable Europe/Anatolia pole of rotation^{3,4} and the locally measured GPS rates around the Marmara Sea¹¹. To fit this data, a 3.0 cm yr^{-1} slip (0.6 cm yr^{-1} on the SNAF, 2.4 cm yr^{-1} on the NNAF) is needed. We reduced slightly the component of opening on the NNAF consistent with the pole to fit local GPS vectors. The depth used was 15 km. In fact, reasonable fits with local GPS could be obtained with locking depths of 10–20 km (ref. 23). However, a locking depth of 10 km seems too small to generate an earthquake of $M = 7.4$, while a locking depth of 20 km requires the slip to be increased to nearly 3.5 cm yr^{-1} , which seems too high.

Fault parameters for earthquakes from 1700 to 1900

All of the $M > 6$ earthquakes in Turkey produce surface faulting that accumulates to produce recognisable morphological features. These allow fault geometry to be mapped, each fault segment having characteristic parameters (length, strike, dip and rake). When this information is combined with historical data describing damage, seismic sea-waves or liquefaction, all of the parameters of earlier earthquakes can be evaluated⁹. Historical records of fifteen $M > 6$ earthquakes exist in the area around Istanbul for 1700 to 1900 (ref. 7). The catalogue separates events into those with $M > 7$ and those with $6 \leq M < 7$, with $M > 7$ events breaking 2 or more segments while the smaller ones occur on single segments. Segments associated with an event are initially identified mainly from reported damage. Scaling relations¹⁵ are used to determine magnitude and moment. Slip is then distributed between the dislocations representing the segments. The slip direction is taken to be parallel to the relative plate motion³. The final parameters are adjusted using two additional criteria. First, if one event was clearly larger than another, its magnitude should be greater and it should involve a greater length of faulting. Second, unless there is clear evidence that two events occupied the same stretch of fault, they are more likely to have occurred on adjacent segments. An example of the former is provided by the 1719 and 1999 events. Damage in the 1719 and 1999 events occurred in the same places. However, in 1719 there was widespread damage in Istanbul. If such shaking had been repeated in 1999 much greater destruction would have occurred in Istanbul. We therefore conclude that, in addition to the segments that ruptured in August and November 1999, a sub-marine segment closer to Istanbul also ruptured in 1719. An example of the latter is provided by events in 1719, 1754, 1766a and 1766b, which occurred in a westward migrating sequence. These events are placed such that every segment of this stretch of fault moved once. Defining the slip regions for the six largest events is more straightforward than estimating parameters of some smaller events, for which there can be a choice of two or even three possible locations separated by 20 km or more. For example the 1894 event had a damage area similar to the 1999 and 1719 events, but the associated destruction was much less, so its rupture zone should be shorter and reduced to one or two fault segments (see Supplementary Information). However, as these events contribute little to the overall slip, errors in their location have no significant effect on the conclusions we report here.

Received 6 September 1999; accepted 31 January 2000.

1. Barka, A. A. 17 August 1999 Izmit earthquake. *Science* **258**, 1858–1859 (1999).
2. Barka, A. A. The North Anatolian Fault zone. *Ann. Tectonicae* **VI** (suppl.), 164–195 (1992).
3. Reilinger, R. E. *et al.* Global Positioning System measurements of the present-day crustal movements in the Arabia-Africa-Eurasia plate collision zone. *J. Geophys. Res.* **102**, 9983–9999 (1997).
4. Hubert, A. *La Faille Nord-Anatolienne (Cinématique, Morphologie, Localisation, Vitesse et Décalage Total) et Modélisations Utilisant la Contrainte de Coulomb sur Différentes Echelles de Temps*. Thesis, Denis Diderot Univ., Paris VII (1998).
5. Armiño, R., Meyer, B., Hubert, A. & Barka, A. Westward propagation of the North Anatolian fault into the northern Aegean: Timing and kinematics. *Geology* **27**, 267–270 (1999).
6. Ambraseys, N. N. Some characteristic features of the North Anatolian fault zone. *Tectonophysics* **9**, 143–165 (1970).
7. Ambraseys, N. N. & Finkel, C. Long-term seismicity of Istanbul and of the Marmara Sea region. *Terra Nova* **3**, 527–539 (1991).
8. King, G. C. P., Stein, R. S. & Lin, J. Static stress changes and the triggering of earthquake. *Bull. Seismol. Soc. Am.* **84**, 935–953 (1994).
9. Nalbant, S. S., Hubert, A. & King, G. C. P. Stress coupling between earthquakes in northwest Turkey and the north Aegean Sea. *J. Geophys. Res.* **103**, 24469–24466 (1998).
10. Stein, R. S., Barka, A. A. & Dieterich, J. H. Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering. *Geophys. J. Int.* **128**, 594–604 (1997).
11. Straub, C., Khale, H. G. & Schindler, C. GPS and geological estimates of the tectonic activity in the Marmara Sea region, NW Anatolia. *J. Geophys. Res.* **102**, 27587–27601 (1997).
12. Barka, A. A. & Kadinsky-Cade, K. Strike-slip fault geometry in Turkey and its influence on earthquake activity. *Tectonophysics* **7**, 663–684 (1988).
13. Harris, R. A. Stress triggers, stress shadows and implications for seismic hazard. *J. Geophys. Res.* **103**, 24347–24358 (1998).
14. Harris, R. A. & Simpson, R. W. In the shadow of 1857: Effect of the great Ft. Tejon earthquake on subsequent earthquakes in southern California. *Geophys. Res. Lett.* **23**, 229–232 (1996).
15. Kanamori, H. & Anderson, D. L. Theoretical basis of some empirical relations in seismology. *Bull. Seismol. Soc. Am.* **65**, 1073–1095 (1975).
16. Wittlinger, G. *et al.* Tomographic evidence for localized lithospheric shear along the Altyn Tagh fault. *Science* **282**, 74–76 (1998).

17. Pollitz, F. F. & Dixon, T. H. GPS measurements across the northern Caribbean plate boundary zone: impact of postseismic relaxation following historic earthquakes. *Geophys. Res. Lett.* **25**, 2233–2236 (1998).
18. Ambraseys, N. N. & Finkel, C. *Historical Seismograms and Earthquakes of the World* (eds Lee, W. H. K., Meyers, H. & Shimazaki, K.), 173–180 (Academic, San Diego, 1988).
19. Dieterich, J. A constitutive law for rate of earthquake production and its application to earthquake clustering. *J. Geophys. Res.* **99**, 2601–2618 (1994).
20. Sornette, A. & Sornette, D. Self organized criticality and earthquake. *Europhys. Lett.* **9**, 197 (1989).
21. Bowman, D. D. *et al.* An observational test of the critical earthquake concept. *J. Geophys. Res.* **103**, 24359–24372 (1998).
22. Okada, Y. Internal deformation due to shear and tensile fault in a half-space. *Bull. Seismol. Soc. Am.* **82**, 1018–1040 (1982).
23. Segall, P. & Harris, R. Slip deficit on the San Andreas Fault at Parkfield, California, as revealed by inversion of geodetic data. *Science* **233**, 1409–1413 (1986).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank F. F. Pollitz and R. Harris for comments on the manuscript. This work was supported by the EEC FAUST programme and the INSU-CNRS PNRN programmes.

Correspondence and requests for materials should be addressed to A.H.-F. (e-mail: ferrari@princeton.edu).

Fine-scale heterogeneity in the Earth's inner core

John E. Vidale*† & Paul S. Earle*

Earth and Space Science Department* and Institute of Geophysics and Planetary Physics†, University of California at Los Angeles, Los Angeles, California 90095-1567, USA

The seismological properties of the Earth's inner core have become of particular interest as we understand more about its composition and thermal state^{1,2}. Observations of anisotropy and velocity heterogeneity in the inner core are beginning to reveal how it has grown and whether it convects^{3,4}. The attenuation of seismic waves in the inner core is strong, and studies of seismic body waves^{5,6} have found that this high attenuation is consistent with either scattering or intrinsic attenuation⁵. The outermost portion of the inner core has been inferred to possess layering and to be less anisotropic than at greater depths^{7–10}. Here we present observations of seismic waves scattered in the inner core which follow the expected arrival time of the body-wave reflection from the inner-core boundary. The amplitude of these scattered waves can be explained by stiffness variations of 1.2% with a scale length

Table 1 Earthquakes and explosions in this study

Date dd/mm/yr	Time	Latitude (°N)	Longitude (°E)	Depth (km)	Distance
28/02/69	04:25:36.9	36.23	–10.48	33	68°
31/05/70	20:23:27.3	–9.18	–78.82	33	68°
02/08/71	07:24:56.0	41.37	143.44	45	73°
05/09/71	18:35:26.8	46.54	141.15	14	70°
06/09/71	13:37:10.1	46.76	141.39	21	70°
09/09/71	23:02:06.8	44.34	150.85	7	69°
27/09/71	05:59:55.8	73.39	54.92	0	59°
24/11/71	19:35:28.5	52.85	159.22	99	57°
28/02/73	06:37:54.4	50.51	156.58	62	60°
17/06/73	20:37:52.1	42.65	146.08	11	70°
24/06/73	02:43:22.8	43.29	146.43	26	70°
27/10/73	06:59:58.0	70.80	53.96	0	62°
15/05/74	18:59:56.1	49.98	156.22	58	60°
29/08/74	09:59:56.2	73.39	54.91	0	59°
09/10/74	07:32:00.6	44.64	150.09	34	67°
02/11/74	04:59:57.4	70.83	53.82	0	62°

of 2 kilometres across the outermost 300 km of the inner core. These variations might be caused by variations in composition, by pods of partial melt in a mostly solid matrix or by variations in the orientation or strength of seismic anisotropy.

Here we examine 12 earthquakes and 4 nuclear explosions in the distance range 58° to 73°. The events, listed in Table 1, were recorded on LASA (Large Aperture Seismic Array) between 1969 and 1975, and were less than 100 km deep. LASA consisted of up to 525 short-period vertical-component seismometers buried to a depth of 70 m and spread across an aperture of 200 km in Montana^{11,12}.

Our seismograms mostly extend from the P wave back past the P'P' arrival, spanning more than half an hour. An arrival that is not explained by previously identified ray-paths appears near the predicted arrival time for PKiKP, which is the faint reflection from the boundary between the inner and outer core. Generally, the boundary between the inner and outer core appears locally flat and sharp^{8,13,14}. Figure 1 shows the onset time, duration, and slowness of the energy incident on LASA in this interval. The image is the logarithmic average of slant stacks of the 16 events generated from the seismograms after they had been passed through a 1-Hz band-pass filter.

We attribute the 200 s of observed ground motion to inner-core scattering (ICS) (Fig. 1). This energy arrives nearly vertically from the direction of the inner core. The energy builds for a few tens of seconds, then gradually fades back into the more omnidirectional background. Most of the energy lies between -0.02 and 0.02 s km⁻¹, consistent with paths from the inner core. The initial tens of seconds of the ICS have an average slowness near 0.01 s km⁻¹, close to that expected for PKiKP. These unexpected arrivals are not visible in the seismograms without stacking. PKiKP is barely visible, as expected at these distances^{8,14,15}.

Also visible in Fig. 1 is a comparable amount of energy with the same slowness as the direct P wave, which also fades with time. This late-arriving P coda is scattered near the event. LASA is sufficiently dense and wide to clearly distinguish between the inner-core arrivals and the P coda. The background noise is generated by the event; it diminishes monotonically with increasing time after the event, and is higher than the pre-event noise.

Our 16 events sample four distinct regions of the inner core (Fig. 2a). The ray paths of PKiKP and the ICS are shown in Fig. 2b. Many of the earthquakes occurred near Japan, and the explosions were in Novaya Zemlya; there were also earthquakes under Peru and

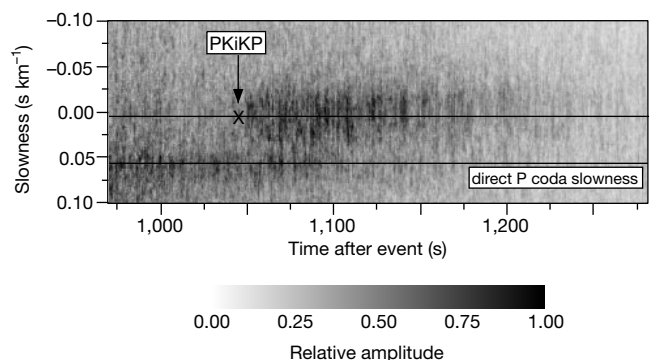


Figure 1 Average seismic-wave amplitude arriving between 970 and 1,280 s after all events. The image shows logarithmically averaged envelopes of the slant stacks for all 12 earthquakes and 4 nuclear explosions. The image shows the time and slowness, which is the reciprocal of apparent velocity, of energy incident on LASA. Three features of the signal are visible: (1) a stripe near 0.06 s km⁻¹ slowness, which corresponds to late P coda and aftershocks, (2) a stripe of energy from 1,050 s until 1,250 s that we interpret as inner-core scattering (ICS), and (3) uniform background source-generated noise that diminishes with time. All subarrays are included in this stack for maximum slowness resolution.

near Iberia. The stack from each event shows ICS energy, although only by stacking all 16 events together do we find the fairly smooth envelope shown in Fig. 3.

We improve the signal-to-noise ratio of the stacks by excluding the outer three of the six rings of the LASA array, at the cost of reducing the high resolution in slowness from inclusion of all rings that is visible in Fig. 1. This winnowing left a 30-km array with more than 120 stations, and resulted in the cleanest stacks with sufficient slowness resolution to separate the ICS from the P coda. The reduced aperture also means that most of the energy incident from the direction of the inner core will stack coherently. In contrast, with the full array, only some of the energy from the inner core is fully in focus at each slowness and back-azimuth.

The amplitude of the ICS averages about 2% of the peak amplitude of PcP for the nuclear tests on 27 September 1971 and 29 August 1974 (Fig. 3). The other two explosions had clipped PcP arrivals, rendering them unsuitable for this calibration. Figure 3 also shows that the ICS lasts far longer than the source of the seismic radiation, which is a brief nuclear explosion. The small amplitude helps explain why ICS has not been previously identified. With an onset that requires 50 s to reach its peak and a signal so small that it cannot be seen in the individual seismograms, only a dense array can provide definitive proof of the strength of ICS. PKiKP is barely visible with amplitude of 3% of PcP, close to its expected value, which is very small because of a small inner-core boundary reflection coefficient^{8,14}.

Figure 4a compares the ICS envelope with the background noise from the inner three rings of LASA stacked for all 16 events. Figure 4b shows the ICS envelope after removing an estimate of uniform background noise. It is clear that the ICS starts near the calculated time for PKiKP, but that it takes 50 s to grow to its peak. PKiKP itself

is only weakly visible for three events, and does not stand out in the stack.

As ICS has not been previously identified, and its presence implies strong fine-scale structure in the inner core, we have extensively explored possible alternative explanations. The high-frequency content indicates incident P waves, the small slowness indicates near-vertical propagation from the direction of the inner core, and the long travel time indicates a long path. We have failed to identify viable alternatives to ICS; PcPPcP (a double reflection from the core–mantle boundary) and P scattering to PKP at the core–mantle boundary come the closest to consistency with the ICS phase.

The calculated envelope of ICS for 1.2% r.m.s. variations in both density and elastic parameters λ and μ , with a correlation distance of 2 km, is shown in Fig. 4b. (Here λ is Lamé's parameter, and μ is the shear modulus.) Our observation geometry requires scattering angles near 90°, which is most sensitive to variations in λ , but variations in μ and density are not well resolved. Our three-dimensional calculation assumes single scattering¹⁶ with equations modified to account for the non-Poisson velocity ratio v_p/v_s of the inner core (v_p and v_s are the velocities of P- and S-waves, respectively). Heterogeneities are parametrized by an exponential autocorrelation function. The elastic structure of the PREM model, with a value of seismic attenuation Q_p of 360 (ref. 6), is assigned throughout the inner core. Without attenuation in the inner core, the ICS would take about 100 s to rise to its peak and last 350 s, until its termination by the reflection from the far side of the inner core. However, the low Q_p in the inner core has the effect that only the scattering from the upper few hundred kilometres of the inner core survives to be seen. So our results only apply to the outer 300 km of the inner core; the deeper scattering structure is unresolved.

Assigning formal error bounds to our model parameters is difficult, given the trade-off between scale-length, r.m.s. variations, inner-core Q , depth distribution of the scatterers, and choice of autocorrelation function. The scale-length of 2 km was chosen because it produces the maximum theoretical amplitude with a minimum r.m.s. variation. Thus, given our scattering model, 1.2% is a lower bound on the r.m.s. variation. Our purpose is to demonstrate that a physically reasonable model exists that fits the data, adding strong support to our hypothesis of fine-scale structure within the inner core.

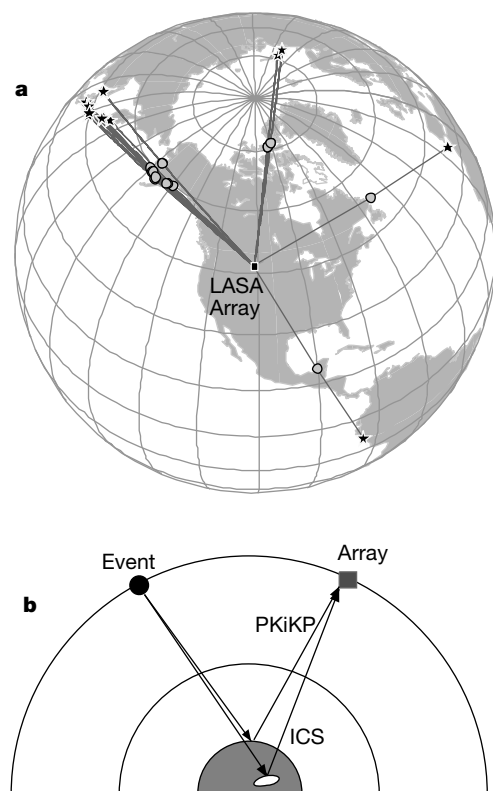


Figure 2 Location map of explosions, earthquakes, LASA, and midpoints of ray-paths. **a**, Map of events (stars), LASA (square), and the surface projection of the PKiKP reflection points (circles) on the inner-core boundary. **b**, Cross-section showing the ray paths of PKiKP and ICS.

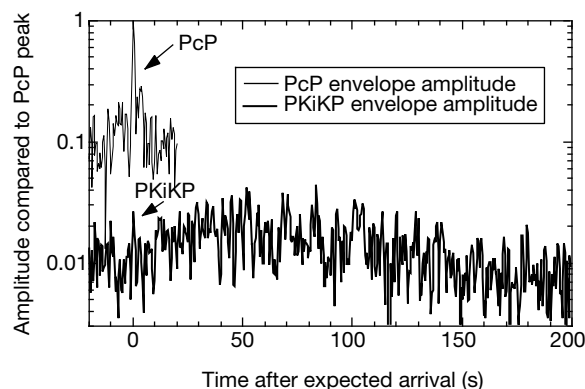


Figure 3 Comparison of ICS with PcP for the 27 September 1971 and 29 August 1974 nuclear tests. Both phases have been aligned with respect to their calculated arrival time. The PcP envelopes are averaged from 0.031 to 0.035 s km⁻¹ and the ICS envelopes are averaged from 0.002 to 0.006 s km⁻¹ to suppress noise. Each event is coherently stacked, then an envelope is taken and the two events are averaged together. The record sections for the two events are very similar, consistent with the reported locations of the explosions being within 400 m.

The Born single-scattering approximation is valid for cases in which the energy loss due to scattering is small. Following Aki and Richards¹⁷ (equation 13.53), and considering only the outer 300 km of the inner core that contributes to ICS waves, we obtain a fractional energy loss of around 10% for our preferred model.

It is difficult to estimate the scattering Q of our inferred heterogeneity in the inner core because we do not well constrain μ and density variations. Q_p at 1 Hz is seen in this study, and in other body-wave work^{5,6}, to be near 300. Normal mode analyses^{18,19} imply higher Q_p , which translates to a less-attenuating core. It is plausible that the attenuation from scattering by fine-scale elastic structure is responsible for the discrepancy, because the normal modes would not be attenuated by such structure.

This magnitude of heterogeneity has several possible explanations. Variation in the orientation of anisotropy may lead to large apparent velocity variations. Some studies predict large variations in the compressional-wave velocity with propagation direction^{20,21}, although other studies predict far less variation²². The work we report here precludes the possibility that the inner core is a single crystal of iron, which would exhibit less scattering²². One-dimensional models of small-scale textural anisotropy show that about 8%

velocity variation on a scale length of 1.2 km can explain the attenuation of PKiKP⁵, although intrinsic attenuation was given as an alternative possibility. The present study requires scattering, and finds that three-dimensional heterogeneity fits ICS much better than one-dimensional layering.

A variable degree of partial melt may also play a role. The measured seismic shear-wave velocities of the core¹⁸ are too slow to match laboratory measurements on pure iron²⁰. This discrepancy may be due either to near-melting softening or to impurities accompanying iron that reduce shear velocity, like sulphur²³. Because the inner core is probably nearly adiabatic, short-wavelength lateral variations in the degree of softening would probably be caused by variations in composition, which could lead to variations in melting point²⁴.

The ICS energy that we report here could be used to determine the rotation rate of the inner core with high precision: this could be achieved if sufficiently dense array recordings of similar sources, repeated at long time intervals, were analysed. This is possible because the scattered-wave travel times are much more sensitive to the movement of heterogeneities than are the transmitted-wave travel times that are currently used to estimate rotation rate²⁵. □

Received 2 August 1999; accepted 25 January 2000.

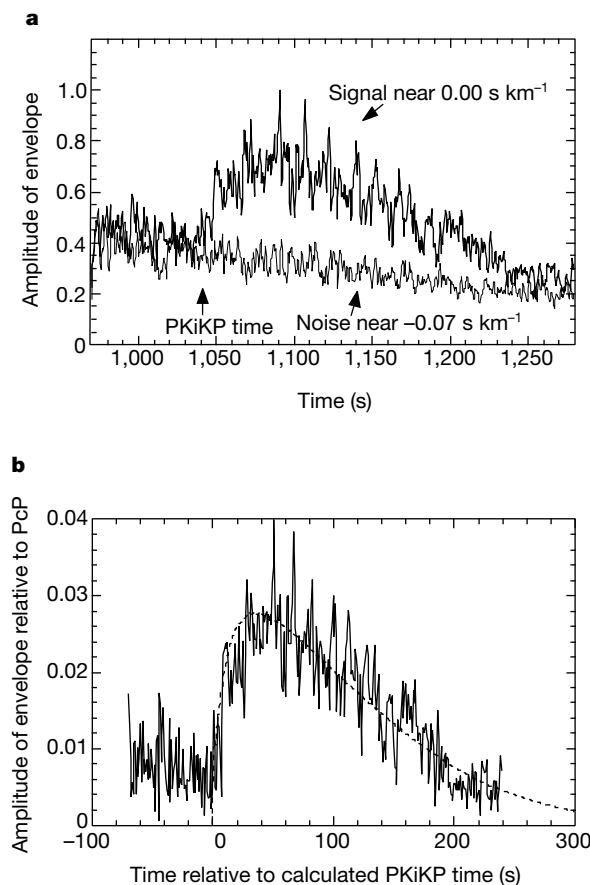


Figure 4 Stacks of inner-core scattering and noise amplitude. Only the three inner subarrays have been included in these stacks. **a**, The inner-core scattering, an average of the ten stacks between -0.005 and $+0.005$ s km⁻¹, is marked by the heavy line. The background noise, an average of the 10 stacks between -0.065 and -0.075 s km⁻¹, where discrete arrivals are not expected or observed, is marked by the thin line. **b**, An estimate of the inner-core scattering after removing noise at the level of the -0.07 s km⁻¹ line in **a**. The noise is removed by squaring the heavy 0.0 s km⁻¹ trace in **a**, subtracting the square of a straight line fit to the noise, and taking the square root. Superimposed is the scattering envelope calculated for our best-fitting model of heterogeneity (dashed line), which has 1.2% variations in the elastic parameters μ and ρ , and density, with 2-km wavelength uniformly distributed across the inner core. PREM elastic velocity structure and a Q of 360 are used in our calculation.

- Anderson, O. L. Mineral physics of iron and of the core. *Rev. Geophys.* **33**, 429–441 (1995).
- Boehler, R. Temperatures in the Earth's core from melting point measurements of iron at high static pressures. *Nature* **363**, 534–536 (1993).
- Tanaka, S. & Hamaguchi, H. Degree one heterogeneity and hemispherical variation in anisotropy in the inner core from PKP(BC) - PKP(DF) times. *J. Geophys. Res.* **102**, 2925–2938 (1997).
- Romanowicz, B., Li, X. D. & Durek, J. Anisotropy in the inner core—could it be due to low-order convection. *Science* **274**, 963–966 (1996).
- Cormier, V. F., Xu, L. & Choy, G. L. Seismic attenuation in the inner core: Viscoelastic or stratigraphic? *Geophys. Res. Lett.* **21**, 4019–4022 (1998).
- Bhattacharyya, J., Shearer, P. M. & Masters, T. G. Inner core attenuation from short period PKP(BC) versus PKP(DF) waveforms. *Geophys. J. Int.* **114**, 1–11 (1993).
- Song, X. & Helmberger, D. V. Seismic evidence for an inner core transition zone. *Science* **282**, 924–927 (1998).
- Souriau, A. & Souriau, M. Ellipticity and density at the inner core boundary from subcritical PKiKP and PcP data. *Geophys. J. Int.* **98**, 39–54 (1989).
- Shearer, P. M. Constraints on inner core anisotropy from PKP(DF) travel times. *J. Geophys. Res.* **99**, 19647–19659 (1994).
- Song, X. & Helmberger, D. V. Depth dependence of anisotropy of Earth's inner core. *J. Geophys. Res.* **100**, 9805–9816 (1995).
- Capon, J. Analysis of Rayleigh-wave multipath at LASA. *Bull. Seismol. Soc. Am.* **60**, 1701–1731 (1970).
- Green, P. E., Frosch, R. A. & Romney, C. F. Principles of an experimental large aperture seismic array (LASA). *Proc. IEEE* **53**, 1821–1833 (1965).
- Engdahl, E. R., Flinn, E. A. & Romney, C. F. Seismic waves reflected from the Earth's inner core. *Nature* **228**, 852–853 (1970).
- Shearer, P. M. & Masters, T. G. The density and shear velocity contrast at the inner core boundary. *Geophys. J. Int.* **102**, 491–498 (1990).
- Buchbinder, G. G. R., Wright, C. & Poupinet, G. Observations of PKiKP at distances less than 110°. *Bull. Seismol. Soc. Am.* **63**, 1699–1707 (1973).
- Wu, R. -S. & Aki, K. Elastic wave scattering by a random medium and the small-scale inhomogeneities in the lithosphere. *J. Geophys. Res.* **90**, 10261–10273 (1985).
- Aki, K. & Richards, P. G. *Quantitative Seismology: Theory and Methods* Vols 1 and 2 (Freeman, San Francisco, 1980).
- Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
- Widmer, R., Masters, T. G. & Gilbert, F. Spherically-symmetric attenuation within the Earth from normal mode data. *Geophys. J. Int.* **104**, 541–553 (1991).
- Mao, H.-k. *et al.* Elasticity and rheology of iron above 220 GPa and the nature of the Earth's inner core. *Nature* **396**, 741–732 (1998).
- Mao, H.-k. *et al.* Correction—Elasticity and rheology of iron above 220 GPa and the nature of the Earth's inner core. *Nature* **399**, 280 (1999).
- Stixrude, L. & Cohen, R. E. High-pressure elasticity of iron and the anisotropy of the Earth's core. *Science* **267**, 1972–1975 (1995).
- Jephcoat, A. & Olson, P. Is the inner core of the Earth pure iron? *Nature* **325**, 332–335 (1987).
- Jeanloz, R. The nature of the Earth's core. *Annu. Rev. Earth Planet. Sci.* **18**, 357–386 (1990).
- Song, X. D. & Richards, P. G. Seismological evidence for differential rotation of the Earth's inner core. *Nature* **382**, 221–224 (1996).

Acknowledgements

We thank the Albuquerque Seismological Laboratory and specifically R. Woodward and H. Bolton for access to LASA data; we also thank P. Shearer, F. Xu, S. Persh and J. Green for comments on the manuscript. L. Knopoff generalized the formula for scattering to non-Poisson solids.

Correspondence and requests for materials should be addressed to J.E.V. (e-mail: vidale@ucla.edu).

The oldest known anthropoid postcranial fossils and the early evolution of higher primates

Daniel L. Gebo*, Marian Dagosto†, K. Christopher Beard‡, Tao Qi§ & Jingwen Wang§

* Department of Anthropology, Northern Illinois University, DeKalb, Illinois 60115, USA

† Department of Cell and Molecular Biology, Northwestern University Medical School, Chicago, Illinois 60611, USA

‡ Section of Vertebrate Paleontology, Carnegie Museum of Natural History, Pittsburgh, Pennsylvania 15213, USA

§ Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, P.O. Box 643, Beijing 100044, People's Republic of China

The middle Eocene primate family Eosimiidae, which is known from sites in central and eastern China^{1,2} and Myanmar³, is central to efforts to reconstruct the origin and early evolution of anthropoid or 'higher' primates (monkeys, apes and humans)^{1–6}. Previous knowledge of eosimiid anatomy has been restricted to the dentition^{1–3,7} and an isolated petrosal bone⁵, and this limited anatomical information has led to conflicting interpretations of early anthropoid phylogeny^{1–6,8,9}. Here we describe foot bones of

Eosimias from the same middle Eocene sites in China that yield abundant dental remains of this primate. Tarsals of *Eosimias* show derived anatomical traits that are otherwise restricted to living and fossil anthropoids. These new fossils substantiate the anthropoid status of *Eosimias* and clarify the phylogenetic position of anthropoids with respect to other major primate clades. Early anthropoids possessed a mosaic of primitive and derived traits in their postcranial skeletons, reflecting their derivation from haplorhine ancestors that retained many prosimian-like features.

Primate tarsal bones are highly diagnostic among mammals^{10–13}, allowing isolated primate tali and calcanei to be readily identified. Our continuing field studies at Chinese fossil sites have yielded numerous jaws and teeth of eosimiid primates and have led to the recovery of both tali and calcanei that can be referred to *Eosimias* (Fig. 1; see Methods). Overall, the tarsals of *Eosimias* resemble those of omomyids (extinct Eocene primates) and small platyrrhines (South American monkeys) more than those of any other primates. Tali of *Eosimias* preserve three characters that have proved useful in distinguishing the tarsals of strepsirhine and haplorhine primates^{14–16}. First, the articular facet for the fibula on the lateral side of the talar body is vertically orientated rather than sloping gently and laterally; second, the groove on the posterior trochlear shelf for the tendon of the flexor muscle (hallucis longus) occurs plantar to the posterior part of the tibiotalar articular surface rather than lateral to it; and third, the posterior trochlear shelf is small rather than large. All of the talar character states present in *Eosimias* resemble those in haplorhines and differ from alternative conditions

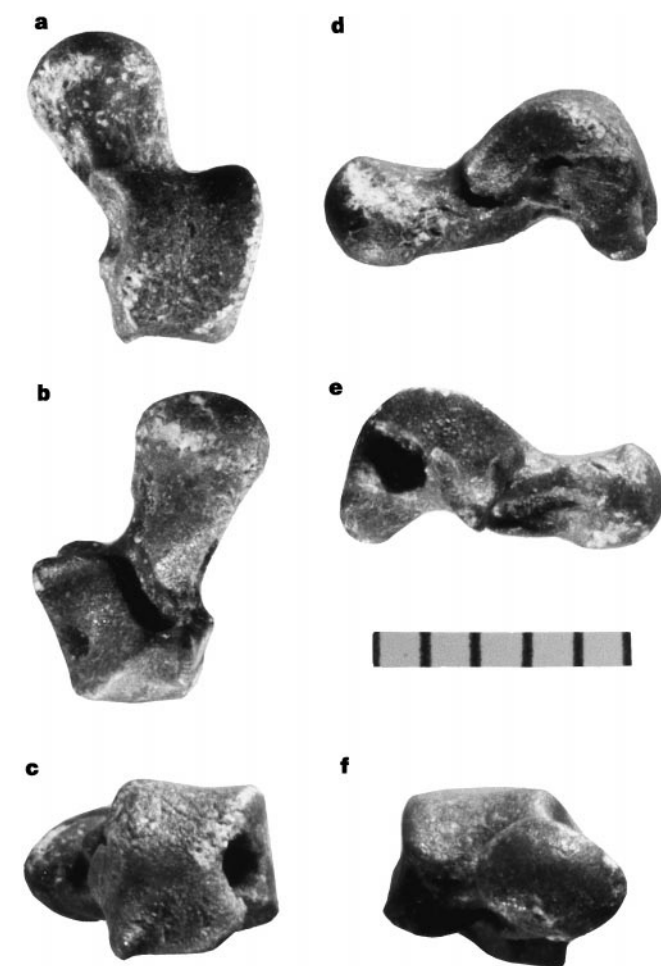


Figure 1 Right *Eosimias* talus (IVPP V11846) from Locality 1, Shanxi Province, China. Views shown are dorsal (a); plantar (b); posterior (c); medial (d); lateral (e) and anterior (f). Scale bar, 5 mm. For talar and calcaneal measurements, see Supplementary Information.

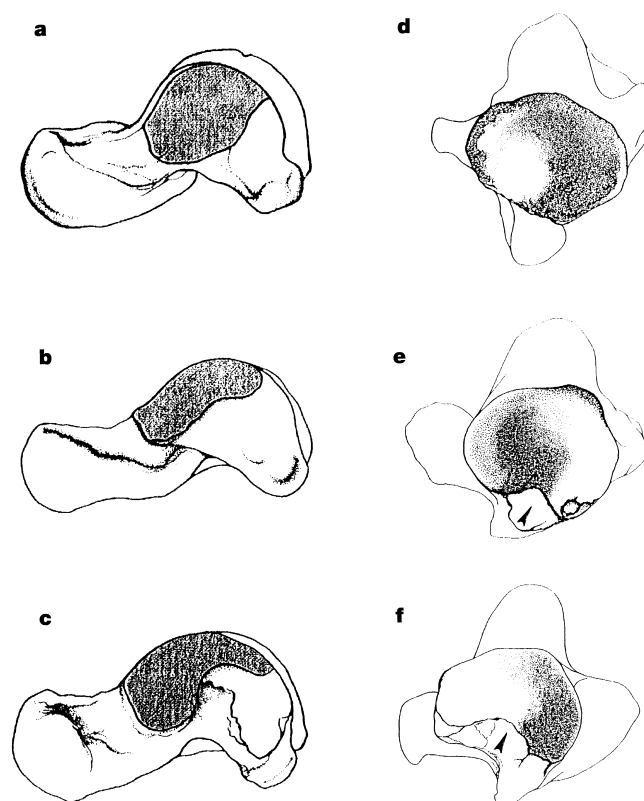


Figure 2 Medial talar and calcaneocuboid features. a–c, Medial views of haplorhine tali. Shading indicates extent of the medial facet. a, *Hemiacodon*, an Eocene omomyid (YPM 24464, Bridger Basin, Wyoming). b, *Eosimias* (IVPP V11849, Shanghuang, Jiangsu Province, China). c, *Saimiri*, a South American monkey (NIU specimen). d–f, Anterior views of calcaneocuboid joints. d, *Hemiacodon*. e, *Eosimias* (IVPP V11848, Shanghuang, Jiangsu Province, China). f, *Saimiri*. Arrows (e, f) point to the nonarticular wedge on the medioplantar surface of the calcaneocuboid joint. Not to scale.

found in strepsirrhines. However, all of these 'haplorhine' character states have been interpreted as primitive for primates^{14–16}. Among haplorhines, eosimiid tali retain some features that are more typical of omomyids than anthropoids (including fossil anthropoids from the Fayum, Egypt). These features include (1) a moderate talar neck angle (30°); (2) a moderately high talar body; (3) a shallow trochlea; (4) a small posterior trochlear shelf; and (5) a relatively narrow talar body. The medial facet on the talar body for the tibial malleolus is reduced (Fig. 2), however, a feature that is diagnostic of anthropoid versus prosimian primates¹⁶. This feature is common in mammals other than primates, and may therefore be interpreted either as a retained primitive character in anthropoids, or as a derived reversal to the primitive condition¹⁵. Optimization of this character by phylogenetic analysis using parsimony (PAUP) identifies this feature as an anthropoid synapomorphy.

Eosimias calcanei can be recognized as belonging to haplorhine primates on the basis of the morphology of the posterior calcaneal facet, which is relatively short and broad, with only a minimal bony distinction between its plantar edge and the supporting bone below it. This morphology is unlike that of adapiform primates, in which the posterior calcaneal facet is ribbon-like (narrow and long), with a well-defined plantar edge. Major primate clades show differential elongation of the anterior or distal part of the calcaneus, which can be quantified by measuring the percentage of total calcaneal length formed by that part of the calcaneus lying distal to the posterior calcaneal facet¹⁷. *Eosimias calcanei* possess a relatively longer distal segment (45–52% of calcaneal length) than do adapiform primates,

being more similar to the squirrel monkey *Saimiri* (range 43–49%; mean 46%) but shorter than most omomyid primates (range 51–56%) (Fig. 3). The calcaneus of *Eosimias* is moderate in its relative transverse width (calcaneal width-to-length ratio, range 38–45%), being narrower than platyrrhines (range 45–63%) but wider than omomyids (range 30–37%). The calcaneocuboid joint of *Eosimias* is also intermediate between those of Eocene prosimians and anthropoids (Fig. 2). The probably primitive condition for primates, exhibited by notharctids and omomyids, is a broad, flat, fan-shaped surface with a centrally located pivot. In contrast, anthropoids possess a joint surface that resembles a circle with a small nonarticular wedge removed from the medioplantar region and a medially shifted pivot. *Eosimias calcanei* possess a relatively flattened joint resembling that of omomyids and notharctids. However, like anthropoids, the joint is rounded, part of the medioplantar section is nonarticular and the pivot is medially shifted.

The reduction of both the medial talotibial facet and the calcaneocuboid joint surface suggests that there was less stability in sustained habitually inverted foot positions. This implies more frequent use of horizontal foot postures and probably horizontal supports in *Eosimias* than in prosimian primates, a functional attribute shared with anthropoids¹⁶. In contrast, the primitive features that *Eosimias* shares with omomyids demonstrate its origin from a leaping ancestry.

A phylogenetic analysis¹⁸ of 11 tarsal characters (Fig. 4; see Methods) shows that a sister group relationship between *Eosimias* and a clade including all other anthropoids is the most parsimonious arrangement of taxa. Bootstrap support is weak because this group is supported by only two tarsal synapomorphies. However, this result agrees with previous analyses of dental and gnathic characters^{1,2}, and bootstrap support for any other grouping (that is, *Tarsius*-anthropoids, adapiforms-anthropoids, omomyids-anthropoids) is much lower (<5%).

The importance of the *Eosimias* tarsals lies in their unique combination of prosimian-like and anthropoid-like traits. This mosaic of primitive and derived characters in *Eosimias*, a basal member of the anthropoid clade, has substantial implications for understanding anthropoid origins. Because the tarsals of *Eosimias* show numerous features that are typical of haplorhines, but lack the derived traits characteristic of strepsirrhines, the oldest known anthropoid postcranial fossils support hypotheses favouring the haplorhine affinities of anthropoids and are inconsistent with hypotheses advocating an adapiform origin for higher primates^{8,19–22}. The precise relationship between early anthropoids and other living and fossil haplorhines is still disputed. The main hypotheses are that (1) anthropoids evolved from Eocene omomyids²³; (2) that anthropoids share more recent common ancestry with living and fossil tarsiers than they do with any other primates^{4,6,24–26}; and (3) that anthropoids diverged from other haplorhines early in the Cenozoic, their most likely sister group consisting of a clade including both tarsiers and omomyids^{1,2,27,28}. Although we favour the latter phylogeny, we acknowledge that tarsal evidence alone cannot discriminate among these options because omomyid tarsal morphology is probably primitive for all haplorhine primates, and would therefore have characterized any pre-tarsioid haplorhine as well. Nevertheless, the eosimiid tarsals described here serve to bridge the wide morphological gap separating prosimian and anthropoid postcrania, providing further insight into the origins of the anthropoid postcranial skeleton. □

Methods

Allocation

IVPP V11846 is the right talus of a small primate from Locality 1 in the Zhaili Member of the Heti Formation, Yuanqu Basin, central China (Fig. 1). After intensive sampling at this locality, only two primate taxa are known to occur here. The larger of these is *Hoanghoniush stehlini*, a basal member of the adapiform family Sivaladapidae that weighed roughly 700 g



Figure 3 Dorsal view of haplorhine calcanei. **a**, *Saimiri* (NIU specimen, a South American monkey). **b**, *Eosimias* (IVPP V11851, Shanghuang, Jiangsu Province, China). **c**, *Hemicodon* (AMNH 12046, an Eocene omomyid). Scale bars, 5 mm.

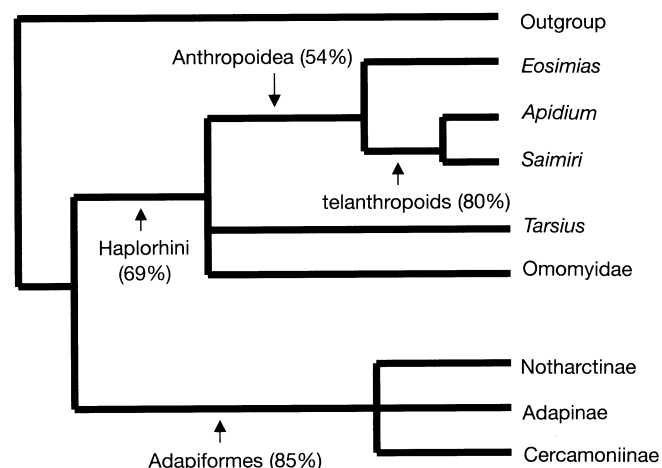


Figure 4 Consensus tree from PAUP analysis. Bootstrap values, given in parentheses, are from 100 replications.

on the basis of regressions of molar area versus body mass in living primates²⁹. The smaller primate from this locality is *Eosimias centennicus*, a basal anthropoid that weighed 90–180 g on the basis of similar regressions² (all estimates of body mass for eosimiids provided here are ranges of mean estimates derived from multiple regressions). Regressions of talar dimensions versus body mass in living primates³⁰ show that IVPP V11846 belonged to a primate weighing 90–147 g. Thus, on the basis of size alone, this talus can be confidently allocated to *Eosimias centennicus*. This allocation permits the recognition of additional *Eosimias* tali ($n = 7$) from the Shanghuang fissures. Estimates of body mass for the Shanghuang sample of *Eosimias* tali range from 57 to 118 g, which overlaps the estimated body mass of *Eosimias sinensis* from Shanghuang on the basis of molar regressions (67–137 g). However, several size classes seem to be represented in the Shanghuang sample of tali, and these probably correspond to more than one eosimiid species. Primate calcanei from Shanghuang ($n = 12$) are referred to *Eosimias* on the basis of their functional congruence with the tali from Shanghuang discussed above, as well as their size, provenance and combination of omomyid-like and anthropoid-like traits.

Phylogenetic analysis

Eleven tarsal characters were subjected to a parsimony analysis using PAUP 4.0 (ref. 18) yielding the nine most parsimonious trees with a consistency index of 0.696. The strict consensus tree is shown in Fig. 4. In all trees, *Eosimias* groups with ‘telanthropoids’²⁵ (in this case, the clade including *Apidium* and *Saimiri*). The ‘outgroup’ consists of character states known in the most likely outgroups of primates: Scandentia, Dermoptera and Plesiadapiformes; these taxa differ insignificantly in the expression of these traits.

Nodes for telanthropoids, Anthropoidea, Haplorhini and Adapiformes are labelled in Fig. 4; the percentages are bootstrap values from 100 replications. Synapomorphies for Anthropoidea are the shape of the calcaneocuboid joint and a reduced medial talar facet. Synapomorphies for telanthropoids are an increased talar neck angle, an increased talar width and loss of the posterior trochlear shelf. Synapomorphies for Haplorhini are an increased distal length of the calcaneus, a relatively short heel, a steep sided talofibular facet with a lantar lip and a centrally located flexor fibularis groove. Synapomorphies for Adapiformes are a sloping talofibular facet, an increase in the size of the posterior trochlear shelf, a long and narrow posterior astragalocalcaneal facet, an increase in the size of the talotibial facet and an increase in the height of the talar body.

Received 31 August 1999; Accepted 26 January 2000.

1. Beard, K. C., Qi, T., Dawson, M. R., Wang, B. & Li, C. A diverse new primate fauna from middle Eocene fissure-fillings in southeastern China. *Nature* **368**, 604–609 (1994).
2. Beard, K. C., Tong, Y., Dawson, M. R., Wang, J. & Huang, X. Earliest complete dentition of an anthropoid primate from the late middle Eocene of Shanxi Province, China. *Science* **272**, 82–85 (1996).
3. Jaeger, J.-J. et al. A new primate from the middle Eocene of Myanmar and the Asian early origin of anthropoids. *Science* **286**, 528–530 (1999).
4. Ross, C., Williams, B. & Kay, R. F. Phylogenetic analysis of anthropoid relationships. *J. Hum. Evol.* **35**, 221–306 (1998).
5. MacPhee, R. D. E., Beard, K. C. & Qi, T. Significance of primate petrosal from middle Eocene fissure-fillings at Shanghuang, Jiangsu Province, People's Republic of China. *J. Hum. Evol.* **29**, 501–514 (1995).
6. Kay, R. F., Ross, C. & Williams, B. A. Anthropoid origins. *Science* **275**, 797–804 (1997).
7. Tong, Y. Middle Eocene small mammals from Liguangzao Basin of Henan Province and Yuanqu Basin of Shanxi Province, central China. *Paleontologia Sinica* **26**, 1–256 (1997).
8. Rasmussen, D. T. in *Anthropoid Origins* (eds Fleagle, J. G. & Kay, R. F.) 335–360 (Plenum, New York, 1994).
9. Godinot, M. in *Anthropoid Origins* (eds Fleagle, J. G. & Kay, R. F.) 235–295 (Plenum, New York, 1994).
10. Lewis, O. J. The joints of the evolving foot. Part III. The fossil evidence. *J. Anat.* **131**, 275–298 (1980).
11. Martin, R. D. *Primate Origins and Evolution: A Phylogenetic Reconstruction* (Princeton Univ. Press, Princeton, 1990).
12. Decker, R. L. & Szalay, F. S. in *Primate Locomotion* (ed. Jenkins, F. A. Jr) 261–291 (Academic, New York, 1974).
13. Dagosto, M. Implications of postcranial evidence for the origin of euprimates. *J. Hum. Evol.* **17**, 35–56 (1988).
14. Beard, K. C., Dagosto, M., Gebo, D. L. & Godinot, M. Interrelationships among primate higher taxa. *Nature* **331**, 712–714 (1988).
15. Dagosto, M. & Gebo, D. L. in *Anthropoid Origins* (eds Fleagle, J. G. & Kay, R. F.) 567–593 (Plenum, New York, 1994).
16. Gebo, D. L. Anthropoid origins—the foot evidence. *J. Hum. Evol.* **15**, 421–430 (1986).
17. Gebo, D. L. Foot morphology and locomotor adaptation in Eocene primates. *Folia Primatol.* **50**, 3–41 (1988).
18. Swofford, D. *Phylogenetic Analysis Using Parsimony* Version 4.0 (Sinauer, Sunderland, Massachusetts, 1998).
19. Simons, E. L. & Rasmussen, D. T. Cranial morphology of *Aegyptopithecus* and *Tarsius* and the question of the tarsier–anthropoid clade. *Am. J. Phys. Anthropol.* **79**, 1–23 (1989).
20. Rasmussen, D. T. The phylogenetic position of *Mahgarita stevensi*: Protoanthropoid or lemuroid? *Int. J. Primatol.* **11**, 439–469 (1990).
21. Simons, E. L. Skulls and anterior teeth of *Catopithecus* (Primates: Anthropoidea) from the Eocene and anthropoid origins. *Science* **268**, 1885–1888 (1995).
22. Gingerich, P. D. in *Evolutionary Biology of the New World Monkeys and Continental Drift* (eds Ciochon, R. L. & Chiarelli, A. B.) 123–138 (Plenum, New York, 1980).
23. Rosenberger, A. L. & Szalay, F. S. in *Evolutionary Biology of New World Monkeys and Continental Drift* (eds Ciochon, R. L. & Chiarelli, A. B.) 139–157 (Plenum, New York, 1980).
24. Cartmill, M. & Kay, R. F. in *Recent Advances in Primatology* Vol. 3 (eds Chivers, D. J. & Joysey, K. A.) 205–213 (Academic, London, 1978).
25. Cartmill, M. in *Evolutionary Biology of the New World Monkeys and Continental Drift* (eds Ciochon, R. L. & Chiarelli, A. B.) 243–274 (Plenum, New York, 1980).
26. Ross, C. in *Anthropoid Origins* (eds Fleagle, J. G. & Kay, R. F.) 469–547 (Plenum, New York, 1994).

27. Beard, K. C., Krishtalka, L. & Stucky, R. K. First skulls of the early Eocene primate *Shoshonius cooperi* and the anthropoid–tarsier dichotomy. *Nature* **349**, 64–67 (1991).
28. Beard, K. C. & MacPhee, R. D. E. in *Anthropoid Origins* (eds Fleagle, J. G. & Kay, R. F.) 55–97 (Plenum, New York, 1994).
29. Fleagle, J. G. *Primate Adaptation and Evolution* 2nd edn (Academic, San Diego, 1999).
30. Dagosto, M. & Terranova, C. J. Estimating the body size of Eocene primates: A comparison of results from dental and postcranial variables. *Int. J. Primatol.* **13**, 307–344 (1992).

Supplementary information is available on Nature's Web site (<http://www.nature.com>) or as paper copy from the London editorial offices of Nature.

Acknowledgements

We thank the Leakey Foundation (D.L.G. and M.D.) and the National Science Foundation (K.C.B.) for their financial support. We also thank our Chinese colleagues, particularly Guo Jianwei, for helping to wash, screen and sort the specimens from the Shanghuang fissures. We also thank Kim Reed-Deemer for her artwork.

Correspondence and requests for materials should be addressed to D.L.G. (e-mail: dgebo@niu.edu).

Soil pathogens and spatial patterns of seedling mortality in a temperate tree

Alissa Packer & Keith Clay

Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

The Janzen–Connell hypothesis^{1,2} proposes that host-specific, distance- and/or density-dependent predators and herbivores maintain high tree diversity in tropical forests. Negative feedback between plant and soil communities could be a more effective mechanism promoting species coexistence because soil pathogens can increase rapidly in the presence of their host³, causing conditions unfavourable for local conspecific recruitment^{4–6}. Here we show that a soil pathogen leads to patterns of seedling mortality in a temperate tree (*Prunus serotina*) as predicted by the Janzen–Connell hypothesis. In the field, the mean distance to parent of seedling cohorts shifted away from maternal trees over a period of 3 years. Seedlings were grown in soil collected 0–5 m or 25–30 m from *Prunus* trees. Sterilization of soil collected beneath trees improved seedling survival relative to unsterilized soil, whereas sterilization of distant soil did not affect survival. *Pythium* spp., isolated from roots of dying seedlings and used to inoculate healthy seedlings, decreased survival by 65% relative to controls. Our results provide the most complete evidence that native pathogens influence tree distributions, as predicted by the Janzen–Connell hypothesis, and suggest that similar ecological mechanisms operate in tropical and temperate forests.

Aggregated spatial distributions of tree species are expected when seed dispersal is highest beneath maternal trees and seedling mortality occurs at random. In contrast, the Janzen–Connell hypothesis predicts that natural enemies will reduce offspring density beneath trees, creating opportunities for heterospecific recruitment. Empirical and theoretical tests of the hypothesis have yielded mixed results^{7–16}. This inconsistency has been attributed, in some cases, to predator satiation, whereby seeds or seedlings close to conspecific trees occasionally escape attack^{17–19}. Escape is less likely with microbial pathogens that exhibit positive density dependence²⁰. Recent studies in temperate communities have suggested that plant–soil feedbacks affect successional dynamics and species diversity^{3,4,21–25}. In this study, we investigated whether negative plant–soil feedback was occurring in *Prunus serotina* (black cherry). Black cherry is a mid-successional tree that produces large numbers of bird-dispersed fruits throughout temperate forests

of eastern North America. Preliminary studies showed that black cherry seedlings experience high mortality in soil collected beneath conspecific adults, but low mortality in soil collected beneath heterospecific adults.

Demographic censuses were conducted beneath six focal black cherry trees from 1996 to 1998. Seed density was significantly

greater 0–10 m from the parent tree than 10–30 m ($187 \pm 14.6 \text{ m}^{-2}$ versus $3.60 \pm 0.11 \text{ m}^{-2}$, $P = 0.05$ (means $\pm$ s.e.)). Although most seedlings germinate within 10 m of the tree, these seedlings have the lowest survival probability (Fig. 1a–c). With random mortality, mean distance to parent will not change over time. However, mean distance to parent of most seedling cohorts shifted outward. For example, the 1996 cohort beneath tree 1 shifted from 11.1 m at germination to 14.3 m 16 months later. Similarly, the 1998 cohort beneath tree 4 shifted from 8.6 m at germination to 10.8 m in just 4 months.

The overall pattern found when seedlings beneath all focal trees were combined revealed that mean distance to parent of survivors was significantly greater (as determined by boot strapped 95% confidence interval, data not shown) than that of germinating seedlings for 1996 (12.4 m versus 14.8 m), 1997 (8.3 m versus 10.0 m) and 1998 (11.0 m compared with 13.4 m) cohorts. There were also pronounced outward shifts between initial and final censuses of saplings established before the study (tree 1, 16.3 m to 16.8 m; tree 4, 12.3 m to 13.2 m; tree 5, 12.2 m to 13.2 m; tree 6, 14.9 m to 16.3 m). Notably, there were no saplings taller than 2 m or with a basal diameter greater than 3 cm in any of the six study arcs, despite high seedling densities. Continued distance-dependent mortality of older individuals could result from current effects of pathogens on sapling root systems or from residual effects of earlier damage.

The change in mean distance to parent differed among trees. Cohorts beneath trees 2 and 3 exhibited a small decrease in mean distance to parent between germination and the final census date, whereas the mean of cohorts beneath trees 1, 4, 5 and 6 increased ($-0.93 \pm 0.44 \text{ m}$ (1 outlier excluded) versus $1.2 \pm 0.44 \text{ m}$, $P = 0.01$). Shifts in distance appear to be correlated with cohort size (trees 2 and 3, 31.8 ± 7.2 seedlings per cohort versus trees 1, 4, 5 and 6, 165.1 ± 51.8 seedlings per cohort, $P = 0.04$). Janzen¹ suggested that trees with small seed crops might have less influence on local seed predators than more productive trees further away. Tree diameter at breast height (trees 2 and 3, $36.8 \pm 8.7 \text{ cm}$, versus trees 1, 4, 5 and 6, $47.6 \pm 7.4 \text{ cm}$, $P = 0.23$) might also be important, although small sample sizes limit statistical power. If tree diameter is correlated with root system size, then smaller trees will have a weaker influence on conspecific seedlings. Finally, both tree and cohort sizes are often related to tree age. With negative feedback, individuals 'culture' (*sensu lato*)²⁵ their soil communities leading to a positive correlation between pathogen accumulation and time that a site has been occupied by a species. Seedlings beneath larger and more fecund trees would therefore experience greater distance- and/or density-dependent mortality.

Increased mortality close to the adult could be due to distance- and/or density-responsive factors. Distance and density are negatively correlated because higher densities occur closer to the tree. We used logistic regression analyses to separate these effects. Data from all cohorts were combined to create one model. Both distance to

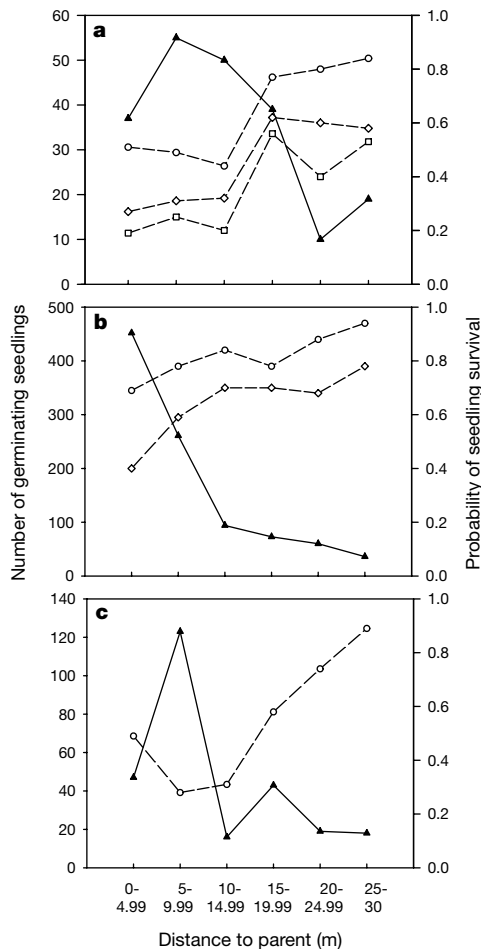


Figure 1 Relationship between distance to parent, initial seedling germination (filled triangles), and probability of seedling survival over time (dashed lines: open circles, after 4 months; open diamonds, after 16 months; open squares, after 28 months). **a**, Spring 1996 cohorts, $n = 212$ seedlings from beneath 3 trees. **b**, Spring 1997 cohorts, $n = 974$ seedlings from beneath 6 trees. **c**, Spring 1998 cohorts, $n = 266$ seedlings from beneath 3 trees.

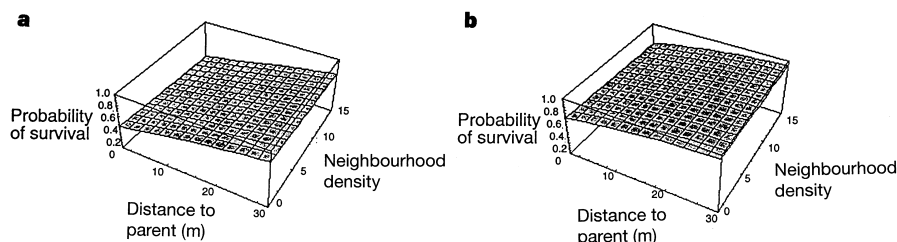


Figure 2 Logistic regression models of the probability of black cherry survival in relation to distance to parent and neighbourhood density. **a**, Model for seedling survival ($n = 1,455$; Likelihood ratio test: $G = 146$, degrees of freedom (d.f.) = 2, $P < 0.00001$). Survival probability is significantly reduced close to the parental tree (Wald statistic = 57.67, d.f. = 1, $P < 0.00001$, $R = 0.17$, Odds ratio = 1.06). Neighbourhood density, although significant in the model, does not strongly influence survival probability (Wald statistic =

45.46, d.f. = 1, $P < 0.00001$, $R = -0.15$, Odds ratio = 0.97). **b**, Model for survival of older individuals ($n = 1,275$; Likelihood ratio test: $G = 68$, d.f. = 2, $P < 0.00001$). Survival probability is significantly reduced close to the tree (Wald statistic = 45.26, d.f. = 1, $P < 0.00001$; $R = 0.19$; Odds ratio = 1.07). Neighbourhood density has no effect on survival (Wald statistic = 0.32, d.f. = 1, $P = 0.57$).

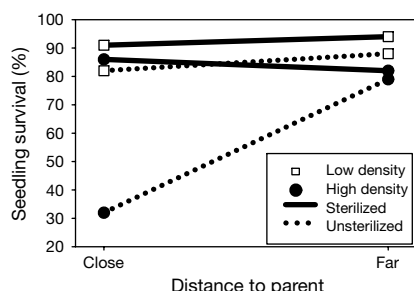


Figure 3 Effect of distance, neighbourhood density and soil sterilization on black cherry seedling survival. In high density treatments, survival was significantly greater after soil collected close to the tree was sterilized. This effect of sterilization was not found with soil collected further from the tree. The data best fitted a logistic regression model that included density, density $\times$ distance, density $\times$ sterilization and distance $\times$ density $\times$ sterilization. Removal of any variable included in the model significantly decreased the model fit (for each variable, $P < 0.0001$).

parent and neighbourhood density at germination were significant predictors of seedling survival, but distance had a stronger effect on survival than density. Survival beneath trees is reduced by 35% relative to survival further away (Fig. 2a). A second model using data from only older individuals also indicated that distance was a better predictor of mortality than density (Fig. 2b). If black cherry root systems support species-specific pathogens, then seedling mortality at a given distance should be relatively constant between years, whereas neighbourhood density may be more variable.

To verify that soil pathogens caused the observed distance- and density-dependent mortality, we conducted a greenhouse experiment using field soil. Independent variables were (1) soil distance (near or far), (2) seedling density (1 or 3 per pot), and (3) soil sterilization (sterilized or not sterilized). There was a three-way interaction between distance, density and sterilization (Fig. 3). In near soil, seedling survival at high density was greatly reduced in unsterilized soil compared with sterilized soil. Seedling survival at high density in far soil was unaffected by sterilization. At low seedling density, survival was unaffected by sterilization, regardless of soil distance. If the pathogen is either at low density or spatially heterogeneous, then the chance of interception by one seedling would be low relative to that of three seedlings, and survival in low density treatments would be high regardless of soil sterilization. Autoclaving soil also kills mycorrhizal fungi and other soil biota, resulting in a nutrient flush. If mycorrhizal fungi enhanced seedling survival beneath the parent tree, sterilized treatments would have reduced seedling survival relative to unsterilized treatments. The opposite effect was observed. Similarly, if a nutrient flush caused the increased survival in sterilized soil, the effect of sterilization would not be distance dependent as we observed. Thus, the results suggest that a soil pathogen causes seedling mortality, especially at near distances and high densities. High seedling density beneath the tree could effectively amplify fungal inoculum such that heavy seed rain would not overwhelm the effects of the pathogen but rather reinforce them, a process opposite to that of predator satiation.

Damping-off type symptoms preceded seedling death in the greenhouse and field. Isolates were obtained from roots of dying seedlings, and the three most common isolates were used to inoculate healthy seedlings. Seedling survival was reduced by an average of 65% in inoculation treatments (Fig. 4). The three isolates were identified as *Pythium* spp. (see Methods) and *Pythium* was later re-isolated from roots of dying seedlings, fulfilling Koch's postulates. Asexual isolates of *Pythium* are difficult to distinguish on the basis of morphological characteristics²⁶ and will require molecular characterization.

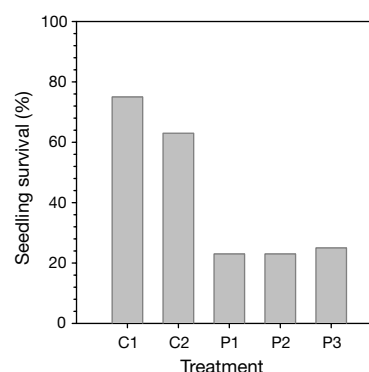


Figure 4 Black cherry seedling survival in control and pathogen inoculation treatments ($n = 40$ per treatment). Control 1, potting mix only. Control 2, 5 ml of sterile nutrient-rich fungal growth medium plus potting mix. P1, P2, P3, 5 ml of appropriate inoculum plus potting mix. Survival was significantly lower in pathogen treatments compared with controls after 19 days ($\chi^2 = 13.8$, d.f. = 4, $P < 0.05$).

Although *Pythium* spp. are often considered generalist pathogens²⁷, they appear to differentially affect seedlings of common tree species. In our preliminary study, black cherry seedling mortality was 100% in black cherry soil, whereas mortality of tulip poplar (*Liriodendron tulipifera*) and dogwood (*Cornus florida*) seedlings was low in black cherry soil. The specificity of the pathogen is further supported by field data suggesting that survival of heterospecific seedlings is not impaired under black cherry. Censuses beneath three focal trees (1, 4 and 5) with high densities of black cherry seedlings revealed only 4 black cherry saplings over 0.5 m high within 10 m (basal stem diameter (b.s.d.) = 18.75 ± 0.75 mm). In contrast, 41 saplings of other species were found within 10 m, including multiple individuals of beech (*Fagus grandifolia*, b.s.d. = 36.56 ± 3.04 mm), sugar maple (*Acer saccharum*, b.s.d. = 47 ± 2.29 mm), dogwood (b.s.d. = 28.71 ± 5.20 mm), ash (*Fraxinus* sp., b.s.d. = 19.67 ± 2.96 mm) and eastern hop hornbeam (*Ostrya virginiana*, b.s.d. = 36 ± 3.00 mm). Seedlings of these species were sparse relative to the cherry seedlings, suggesting that their survival is high beneath black cherry adults. Furthermore, only heterospecific adults occurred within the arcs, suggesting that mortality caused by soil pathogens favours heterospecific recruitment despite the high level of black cherry seed dispersal.

The results of our study on a single species demonstrate that the Janzen–Connell mechanism operates within temperate forests. Tropical predators and herbivores are thought to have greater host-specificity than their temperate counterparts^{1,28} (but see ref. 29); however, less is known about soil microorganisms in either region. An important goal for future research is to determine how frequently the mechanism proposed in the Janzen–Connell hypothesis operates in temperate compared with tropical communities, and the relative importance of animal predators and herbivores versus microbial pathogens. If a smaller proportion of temperate species is damaged by host-specific natural enemies, then similar distance- and density-dependent processes operating in both temperate and tropical communities may lead to different levels of diversity. □

Methods

Demographic censuses

Censuses were conducted beneath 6 black cherry trees at Griffy Lake Nature Preserve, Bloomington, Monroe County, Indiana (3 initiated in 1996 and 3 in 1997). Trees were reproductive and more than 50 m from other conspecific trees. Arcs (30 degrees, extending 30 m) were established with each tree at the apex. Black cherry seedlings and saplings within the arcs were tagged, and their x, y coordinates were recorded. The 1996 and 1997 germinating cohorts were followed for arcs established in 1996, and the 1997 and 1998

germinating cohorts were followed for arcs established in 1997. Individuals established before the beginning of the study ('older') were included in the census. Thus, there were a total of 6 older cohorts and 12 germinating cohorts (3 from 1996, 6 from 1997 and 3 from 1998).

Censuses were conducted in May and September of 1996, 1997 and 1998. Seedlings' distance to parent and distance to all other seedlings were calculated using x, y coordinates. Neighbourhood density was estimated by the number of conspecific seedlings/saplings within a 20 cm radius. Neighbourhood density for individuals along the arc's edge was adjusted by projecting a mirror image of seedling distribution 20 cm within the arc.

Separating distance and density effects

In logistic regression models, distance to parent and neighbourhood density were entered as independent variables, and seedling status (alive or dead) was entered as the dependent variable. Density at germination was used for seedlings, whereas density at the first census was used for older individuals. Logistic regression analyses were conducted using SPSS, version 6.1.3. Comparisons of mean distance to parent at initial and subsequent census dates were made by generating reliability estimates using Resampling Stats, version 3.14.

Effect of soil

Fruits were collected from three trees in Bloomington with large crops. Fruit flesh was removed, and seeds were surface sterilized (1 min 70% alcohol, 3 min 50% bleach, 1 min 70% alcohol, 1 min distilled water). To break dormancy, seeds were stratified in wet sterile sand at 4°C for 5–6 months. Soil was collected at distances of 0–5 m and 25–30 m beneath trees 1, 4 and 6. Soil was sieved and root material was cut into ~1 cm pieces and returned to the soil. Soil samples were diluted 1:1 v/v with sterile potting mix (Metro Mix). One half of the soil in each distance class was autoclaved for 4 h at 211°C. Seedlings that germinated following stratification were planted in each of four soil types (near/far, sterile/unsterile) and watered individually to prevent cross-contamination. Growth and survival of seedlings were monitored for 2 months in the greenhouse. The experiment was conducted in 1998, and again in 1999 with minor modifications. In 1998 all replicates were planted in 4 cm pots, and pots from each soil treatment were placed in the same flat and rotated weekly. In 1999 replicates were planted in randomly distributed 6.5 cm pots. There were no significant differences between years so data from 1998 and 1999 were combined; $n = 34$ per treatment combination except for the treatment combination with both high density and sterilized soil where $n = 28$. Data were analysed using backward conditional logistic regression (SPSS, version 6.1.3).

Pathogen isolation

Upon seedling death in field soil, roots were surface sterilized for 5 min in a 5% bleach solution and rinsed in distilled water. Cross-sections from the leading edge of the disease lesion were plated on corn-meal agar. Isolates were subcultured and maintained on corn-meal agar plates. Isolates were identified to genus by Karen Rane, Diagnostic Technician at Purdue Plant and Pest Diagnostic Laboratory, West Lafayette, Indiana. *Pythium* spp. are in the division Oomycota, kingdom Protocista (not true fungi). Isolates are maintained at Indiana University and available on request. The three most common isolates were used to create inocula grown in a nutrient-rich vermiculite medium. Over 80% of all isolates were one of these three (distinguished on the basis of growth rate; colour and texture were similar). Inoculation treatments are referred to as P1, P2, and P3. To obtain seedlings for screening, ~200 seedlings were collected from an abundant population along a woodland/field margin. Seedlings were within 5 m of each other and in a similar microhabitat, and did not occur beneath a black cherry canopy. Roots were rinsed with distilled water to reduce contamination. Although some seedlings may have been infected when collected, seedlings were randomly distributed among treatments. Plastic pots (6.5 cm) were filled with potting soil and watered. Five millilitres of inoculum were added to a depression created in the centre of each pot. Seedlings were planted into this depression and exposed vermiculite was covered. Two controls were included: potting mix only (control 1); and sterile nutrient-rich medium and potting mix (control 2). There were 40 replicates of each treatment and control. Seedlings were removed upon death. The null hypothesis that survival is independent of treatment was tested using the χ^2 test statistic. Roots from plants dying in each treatment were plated on corn-meal agar to determine whether the fungus could be re-isolated from the roots.

Received 19 November; accepted 22 December 1999.

1. Janzen, D. H. Herbivores and the number of tree species in tropical forests. *Am. Nat.* **104**, 501–508 (1970).
2. Connell, J. H. in *Dynamics in Populations* (eds den Boer, P. J. & Gradwell, G. R.) 298–312 (Center for Agricultural Publishing and Documentation, Wageningen, 1971).
3. Burdon, J. J. *Diseases and Plant Population Biology* (Cambridge Univ. Press, Cambridge, 1987).
4. Florence, R. G. Decline of old-growth forests in relation to some soil microbiological processes. *Ecology* **46**, 52–64 (1965).
5. Bever, J. D., Westover, K. M. & Antonovics, J. Incorporating the soil community into plant population dynamics: the utility of the feedback approach. *J. Ecol.* **85**, 561–573 (1997).
6. Mills, K. E. & Bever, J. D. Maintenance of diversity within plant communities: soil pathogens as agents of negative feedback. *Ecology* **79**, 1595–1601 (1998).
7. Hubbell, S. P. Seed predation and the coexistence of tree species in tropical forests. *Oikos* **35**, 214–229 (1980).
8. Augspurger, C. K. Offspring recruitment around tropical trees: changes in cohort distance with time. *Oikos* **40**, 189–196 (1983).
9. Augspurger, C. K. & Kelly, C. K. Pathogen mortality of tropical tree seedlings: experimental studies

of the effects of dispersal distance, seedling density, and light conditions. *Oecologia* **61**, 211–217 (1984).

10. Clark, D. A. & Clark, D. B. Spacing dynamics of a tropical rain forest tree: evaluation of the Janzen–Connell model. *Am. Nat.* **124**, 769–788 (1984).
11. Connell, J. H., Tracey, J. G. & Webb, L. J. Compensatory recruitment, growth and mortality as factors maintaining rain forest tree diversity. *Ecol. Monogr.* **54**, 141–164 (1984).
12. Becker, P., Lee, L. W., Rothman, E. D. & Hamilton, W. D. Seed predation and the coexistence of tree species: Hubbell's models revisited. *Oikos* **44**, 382–390 (1985).
13. Condit, R., Hubbell, S. P. & Foster, R. B. Recruitment of conspecific adults and the maintenance of tree and shrub diversity in a neotropical forest. *Am. Nat.* **140**, 261–286 (1992).
14. Gilbert, G. S., Hubbell, S. P. & Foster, R. B. Density and distance to adult effects of a canker disease of trees in a moist tropical forest. *Oecologia* **98**, 100–108 (1994).
15. Notman, E., Gorchov, D. L. & Cornejo, F. Effect of distance, aggregation, and habitat on levels of seed predation for two mammal-dispersed neotropical rain forest tree species. *Oecologia* **106**, 221–227 (1996).
16. Willis, C., Condit, R., Foster, R. B. & Hubbell, S. P. Strong density- and diversity-related effects help to maintain species diversity in a neotropical forest. *Proc. Natl Acad. Sci.* **94**, 1252–1257 (1997).
17. Schupp, E. W. The Janzen–Connell model for tropical tree diversity: population implications and the importance of spatial scale. *Am. Nat.* **140**, 526–530 (1992).
18. Burkey, T. V. Tropical tree species diversity: A test of the Janzen–Connell model. *Oecologia* **97**, 533–540 (1994).
19. Cintra, R. A test of the Janzen–Connell model with two common tree species in Amazonian forest. *J. Trop. Ecol.* **13**, 641–658 (1997).
20. Burdon, J. J. & Chilvers, G. A. Host density as a factor in plant disease ecology. *Annu. Rev. Phytopathol.* **20**, 143–166 (1982).
21. Fox, J. F. Alternation and coexistence of tree species. *Am. Nat.* **111**, 69–89 (1977).
22. Woods, K. D. Reciprocal replacement and the maintenance of codominance in a beech maple forest. *Oikos* **33**, 31–39 (1979).
23. Van der Putten, W. H. & Peters, B. A. M. How soil-borne pathogens may affect plant competition. *Ecology* **78**, 1785–1795 (1997).
24. Van der Putten, W. H., Van Dijk, C. & Peters, B. A. M. Plant-specific soil-borne diseases contribute to succession in foredune vegetation. *Nature* **362**, 53–56 (1993).
25. Bever, J. D. Feedback between plants and their soil communities in an old field community. *Ecology* **75**, 1965–1977 (1994).
26. Chen, W., Schneider, R. W. & Hoy, J. W. Taxonomic and phylogenetic analyses of ten *Pythium* species using isozyme polymorphisms. *Phytopathology* **82**, 1234–1244 (1992).
27. Carlisle, M. J. & Watkinson, S. C. *The Fungi* (Academic, Harcourt Brace & Co., New York, 1994).
28. Barone, J. A. Host-specificity of folivorous insects in a moist tropical forest. *J. Animal Ecol.* **67**, 400–409 (1998).
29. May, R. M. How many species are there on Earth? *Science* **241**, 1441–1449 (1988).

Acknowledgements

We thank C. Augspurger, G. Gilbert, J. Holah, P. Kover, C. Lively and J. Price for critical review of the manuscript. We also thank R. Wagner for assistance with fungal isolation and identification, and J. Berg, M. Hallinan, M. Hardesty, T. Levitt, A. Overgaard, M. Selby, L. Somers and members of the Clay lab for field assistance. This work was supported by grants to A.P. from the Indiana Academy of Sciences and Indiana University.

Correspondence should be addressed to A.P. (e-mail: apacker@indiana.edu).

The ecological cost of sex

C. Patrick Doncaster*, Graeme E. Pound*† & Simon J. Cox†

* Division of Biodiversity and Ecology, School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, UK
† Department of Electronics and Computer Science, University of Southampton, Highfield, Southampton SO17 1BJ, UK

Why sex prevails in nature remains one of the great puzzles of evolution^{1,2}. Sexual reproduction has an immediate cost relative to asexual reproduction, as males only express their contribution to population growth through females. With no males to sustain, an asexual mutant can double its relative representation in the population in successive generations. This is the widely accepted 'twofold cost of males'^{1–3}. Many studies^{4–7} have attempted to explain how sex can recoup this cost from fitness benefits associated with the recombination of parental genotypes, but these require complex biological environments that cycle over evolutionary timescales. In contrast, we have considered the ecological dynamics that govern asexual invasion. Here we show the existence of a threshold growth rate for the sexual population, above

which the invasion is halted by intraspecific competition. The asexual population then exerts a weaker inhibitory effect on the carrying capacity of the sexual population than on its own carrying capacity. The stable outcome of this is coexistence on a depleted resource base. Under these ecological circumstances, longer-term benefits of sex may eventually drive out the asexual competitor.

An allele that confers a twofold advantage in growth rate per capita doubles its representation in the population each generation. An asexual species with the same advantage over a sexual species only does so if the population is unbounded by carrying capacity. Here we use a classical ecological model to show this distinction with the fewest possible assumptions. Our model provides a framework for understanding sex and asex as alternative strategies for reproductively isolated sub-populations that compete to consume a common resource. We begin by modelling the dynamics of exploitation competition between two populations of resource consumers. Each has a genetically coded rate constant for conversion of resource uptake into new consumer biomass, and an average lifespan for its members that may be influenced by its environment. We then explore the implications for coexistence if one of these types reproduces sexually and the other invades as an asexual mutant.

Consider a renewing stock of prey exploited to some fraction S of their carrying capacity by a population of predators present at a fraction N of their carrying capacity. Let the predator have a rate of prey consumption per capita that depends directly on prey density. We then represent the dynamics of prey renewal and consumption with a continuous rate equation:

$$\dot{S} = (1 - S)S - NS \quad (1)$$

The first term describes a logistic renewal of prey, from a unitary intrinsic rate to zero at carrying capacity $S = 1$ (in the absence of predation). The second term describes the opposing rate of loss of stock to the predator.

We now distinguish between two species of predator competing to exploit the resource, present at fractions N_1 and N_2 of their respective carrying capacities. Let the rate of consumption by each predator 1 have a component α_{21} accounting for the prey eaten by predator 1 that would otherwise have been eaten by predator 2. In effect, α_{21} describes the consequences to predator 2 of exploitation by predator 1. A matching component α_{12} describes the inhibitory effect of predator 2 on predator 1. If $\alpha_{ij} = 1$, then predator j takes each prey from predator i at the same unitary rate as predator i takes each prey for itself. In this case, the feeding niche of predator j wholly encompasses that of predator i . If $\alpha_{ij} = \alpha_{ji} = 0$, then neither predator uses up any of the prey exploited by the other. In this case, the feeding niches of the predators do not overlap. The conceptual scheme is illustrated in Fig. 1. Two new rate equations then describe the resource available to each species, S_1 and S_2 , as fractions of their respective carrying capacities within the niches of N_1 and N_2 :

$$\dot{S}_1 = (1 - S_1)S_1 - (N_1 + \alpha_{12}N_2)S_1 \quad (2a)$$

$$\dot{S}_2 = (1 - S_2)S_2 - (N_2 + \alpha_{21}N_1)S_2 \quad (2b)$$

Although these equations describe a single prey species that is shared between two predators, so that S_1 and S_2 are not independent, similar conclusions can be drawn from a model with two independent prey species in which $S_1 + S_2 = S$. Setting the rates of change to zero in equations (2) and solving for S_i yields the equilibrium stock of prey available to each predator species:

$$S_1^* = 1 - (N_1 + \alpha_{12}N_2) \quad (3a)$$

$$S_2^* = 1 - (N_2 + \alpha_{21}N_1) \quad (3b)$$

Finally, consider the dynamics of the two predators. Let each predator i recruit in proportion r_i to its exploitation per capita of S_i

prey, and die at a constant rate per capita d_i . Assuming prey renews at relatively fast timescales relative to predator turnover, we can use equations (3) to obtain predator rate equations at prey equilibrium:

$$\dot{N}_1 = r_1[1 - (N_1 + \alpha_{12}N_2)]N_1 - d_1N_1 \quad (4a)$$

$$\dot{N}_2 = r_2[1 - (N_2 + \alpha_{21}N_1)]N_2 - d_2N_2 \quad (4b)$$

Equations (4) describe a logistic recruitment of predator i for a given N_j , from an intrinsic rate per capita r_i to zero at the maximum density of 1 in the absence of losses d_i . This is the classical Lotka–Volterra model of interspecific competition^{8,9}. Crucially, for the purposes of our argument, we have expanded it to allow for separate components of mortality d_i and therefore separate birth and death processes, rather than combining them in the conventional net growth rate. Each predator has a steady-state abundance given by setting its rate equation to zero and solving for N_i :

$$N_1^* = k_1 - \alpha_{12}N_2, \text{ where } k_1 = 1 - \frac{d_1}{r_1} \quad (5a)$$

$$N_2^* = k_2 - \alpha_{21}N_1, \text{ where } k_2 = 1 - \frac{d_2}{r_2} \quad (5b)$$

The constant k_i is the carrying capacity of predator i in the absence of the other consumer. The equilibrium abundance of each thus depends on the abundance of the other. By plotting these zero isoclines on the phase plane of N_1 and N_2 , we can explore the parameter values that allow coexistence of the two predators. In fact, coexistence requires that $k_1/\alpha_{12} > k_2$ and $k_2/\alpha_{21} > k_1$, as illustrated by the example in Fig. 2a. This yields a stable equilibrium of positive (N_1^*, N_2^*) at the intersection of the isoclines, towards which any positive (N_1, N_2) converges. In biological terms, this means that each predator must have a smaller competitive effect on the carrying capacity of the other than it has on its own carrying capacity through intraspecific competition. Any other arrangement of isoclines produces extinction of one or the other predator¹⁰. If predator j exerts a stronger inhibitory effect on the carrying capacity of predator i than it exerts on its own carrying capacity ($k_i/\alpha_{ij} < k_j$), then predator j will replace predator i .

We now consider the particular case of a sexually reproducing

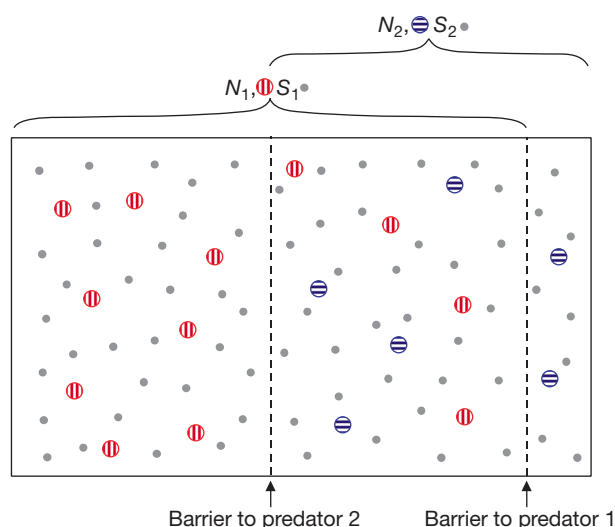


Figure 1 Conceptual scheme for exploitation by a predator 1 (red vertical stripe) and a predator 2 (blue horizontal stripe) of a renewing stock of limiting prey (grey filled dots). Coexistence is possible if the two resource niches are not identical, otherwise the species with the faster potential for population growth (given by r_i/d_i) outcompetes the other at equilibrium resource use, following Gause's competitive exclusion principle. In this example, only the middle section of the arena is exploited by both species ($\alpha_{12} = 0.5$, $\alpha_{21} = 0.8$).

anisogamous predator 1 that produces an asexual mutant predator 2. The cost of males gives predator 1 an intrinsic growth rate per capita half that of predator 2 on average, so $r_1 = 0.5r_2$. This means that we assume an identical intrinsic birth rate per capita for asexual individuals and sexual females, and zero birth rate for males which constitute half the sexual population. If the two types differ only in this respect, we can expect $d_1 = d_2$. It is then useful to define $R_0 = r_1/d_1$, which describes the capacity for growth of the sexual population. R_0 refers to the average number of offspring born to each sexual individual before it dies, in the absence of both intra- and interspecific competition (the subscript zero referring to the sexual population at time zero when the first consumer converts prey into offspring). The carrying capacities now become:

$$k_1 = 1 - \frac{1}{R_0} \text{ and } k_2 = 1 - \frac{1}{2R_0} \quad (6)$$

The twofold difference in growth capacity yields a larger carrying capacity for predator 2 than predator 1, reflecting its greater efficiency of converting food into new predator biomass. This is accommodated by a reduction in equilibrium prey stock. Figure 2b shows the sequence of events as an asexual mutant invades a sexual population. The larger carrying capacity of the asexuals contributes to their higher equilibrium presence, at the expense of the sexuals. Crucially, however, the twofold greater growth capacity of the asexuals has an influence on the dynamics of asexual invasion that depends on the absolute size of R_0 . Larger values of R_0 reduce

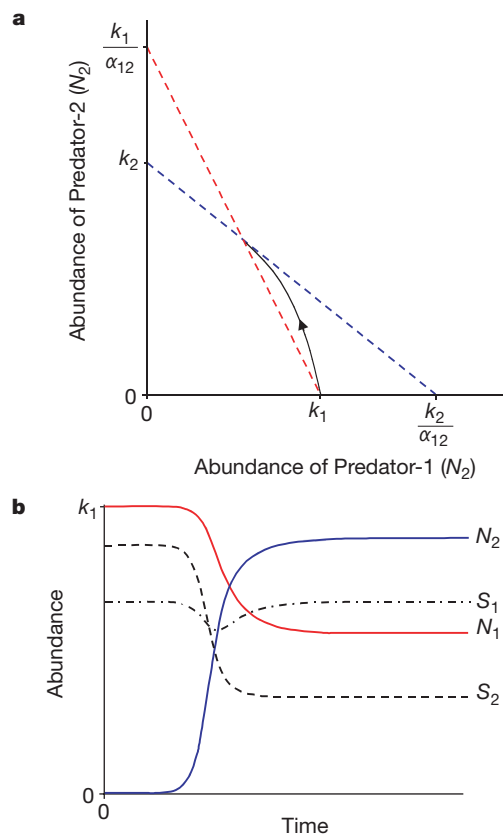


Figure 2 Population dynamics of exploitation competition. **a**, Phase plane for two species of competing consumers. The steeper of the two broken lines is the isocline for exploitation by predator 1 at a given abundance of predator 2, and the other is the isocline for exploitation by predator 2 at a given abundance of predator 1 (equations (5)). The example shows stable coexistence at the intersection of the isoclines when a sexual predator 1 is invaded by an asexual predator 2 with twice the capacity for growth (arrow). $\alpha_{12} = 0.5$, $\alpha_{21} = 0.8$, $r_1 = 0.5r_2$, $d_1 = d_2 = 0.04$, $R_0 = 2.5$ (predator 1). At $t = 0$, $N_1 = k_1$, $N_2 = 0.0001$. Abundance and time are non-dimensionalized. **b**, Re-equilibration over time of example in **a**. Each predator i uses resource S_i , and $\dot{N}_i = r_i S_i N_i - d_i N_i$, with $\dot{S}_i$ given by equations (2).

the difference between k_2 and k_1 (equations (6)), and therefore between N_2^* and N_1^* (equations (5)). If R_0 is substantially greater than unity, the twofold cost of males will have little consequence on the dynamics of an asexual invasion. Figure 3 shows how sexuals and asexuals reach an asymptotic ratio as R_0 becomes infinitely large. Both types then have an infinite lifespan ($d_1 = d_2 = 0$), which means that their equilibrium populations have zero birth rates and the asexuals are thus unable to realize their superior growth capacity. The ratio of sexuals to asexuals diminishes as R_0 approaches unity. Sexuals cannot persist with asexuals below a threshold R_0 , defined by the twofold difference in reproductive capacity and the competition coefficient of the asexual α_{12} (Fig. 3). Figure 4 shows how a high growth capacity permits coexistence even with a high competitive impact of asexuals on sexuals (α_{12} approaching 1), and that coexistence at lower R_0 requires a correspondingly lower minimum α_{12} .

Provided that $k_1/k_2 > \alpha_{12}$, the sexual population coexists with the asexual invaders, or even drives them out altogether if $k_1 > k_2/\alpha_{21}$ (Fig. 2a). We assume that the sexual predator 1 has less impact on the growth of the asexual predator 2 than the asexual has on its own growth, so $\alpha_{21} < 1$. This is because the self-identical asexual individuals (barring further mutations) must experience the strongest possible intraspecific competition. We further propose that the asexual predator 2 is likely to have less impact on the sexual predator 1 than the sexual has on its own growth, even at the very start of the invasion, so $\alpha_{12} < 1$. This is reasonable if the asexual mutant has been cloned from one or a few sexual genotypes, so that its population represents only a small sample of the genetic pool to which the sexual population belongs¹¹. This type of founder effect restricts the asexual to a narrower niche than is occupied by the sexual, given genetic variability for resource exploitation in the sexual population¹². Recent examples of resource partitioning between clones support a relationship between environmental heterogeneity and genetic variation^{13–16}. Empirical evidence that differences in niche breadth may explain coexistence of sexual and asexual populations has been found for several species, including the freshwater snail *Potamopyrgus*¹⁷, the fish *Poeciliopsis*¹⁸ and the lizard *Cnemidophorus*¹⁹. In the longer term, character displacement in the sexual predator 1 and mutation and selection in the asexual predator 2 seem likely to further reduce the value of α_{12} . In the event that $d_1 < d_2$ such that $k_1 > k_2$, the sexual type persists even with $\alpha_{12} = 1$. Disruptive selection is likely to cause d_1 to change with respect to d_2 , particularly in the form of character displacement by the genetically variable sexuals as a result of competition with the self-identical asexuals²⁰.

Although the cost of males is clearly ecology dependent, it remains widely cited as a twofold disadvantage that must be

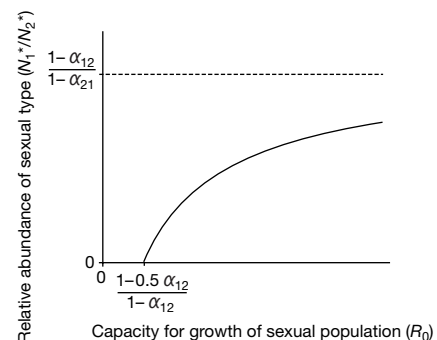


Figure 3 Influence of the growth capacity of sexual predator 1 ($R_0 = r_1/d_1$) on the equilibrium ratio of sexuals to asexuals. As in Fig. 2, $r_1 = 0.5r_2$, $d_1 = d_2 = 0.04$. The relative abundance of sexuals has an asymptote for large R_0 , given by the dotted line. Smaller values of α_{12} increase this asymptote, and decrease the threshold R_0 for coexistence. Unless $\alpha_{12} \geq 1$, there always exists some R_0 that will support a sexual population.

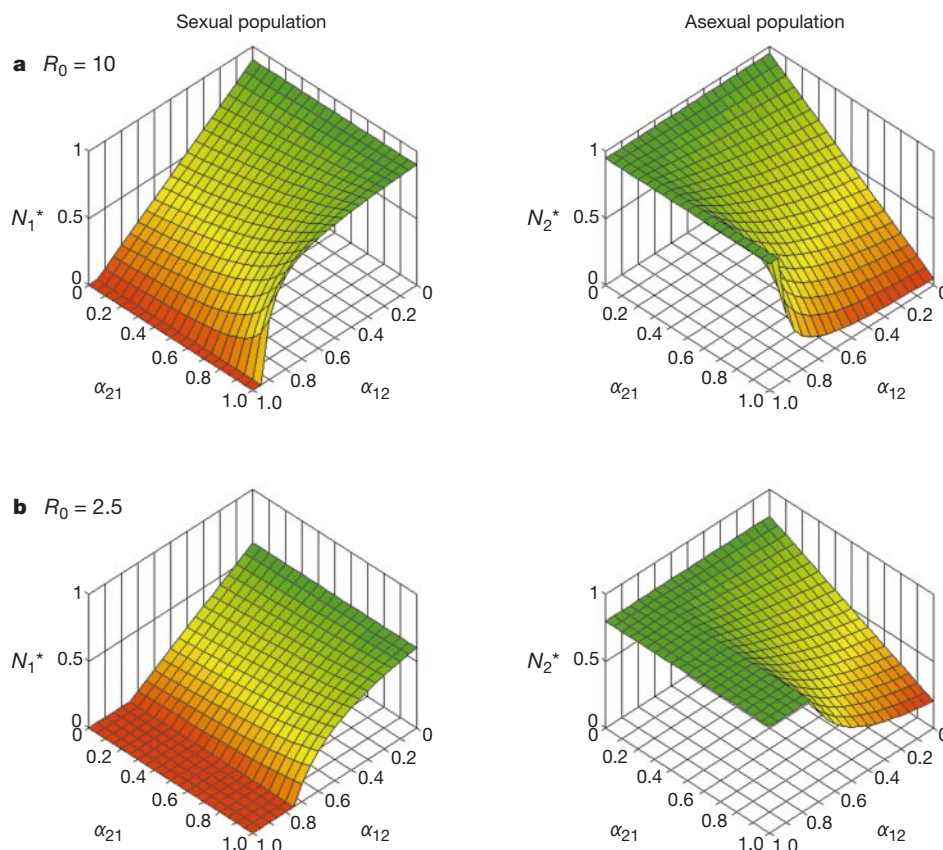


Figure 4 Influence of competition coefficients α_{12} and α_{21} on equilibrium densities of sexual and asexual populations. **a**, Sexual population with a high capacity for growth,

$R_0 = 10$, given by $d_1 = 0.01$. **b**, Sexual population with a lower capacity for growth, $R_0 = 2.5$ as Fig. 2, given by $d_1 = 0.04$; other parameter values as in Figs 1–3.

recouped in adaptive payoffs from the recombination of parental genotypes^{2,21–25}. Why do evolutionary biologists continue to ignore the evidence from population ecology? We note from the literature that ecological outcomes have previously been proposed with the aid of purpose-built models rather than drawing on the standard framework of Lotka–Volterra competition equations, with some loss of generality in consequence. A pioneering study by Case and Taper²⁶ sought conditions for coexistence from a more elaborate treatment incorporating marriage functions and resource-dependent death rates. They had dismissed the classical Lotka–Volterra model as unrepresentative of competition between sexual and asexual types, on the grounds that it failed to separate birth and death processes in the predator dynamics. Indeed, the usual quotations of equations (4) in the literature omit explicit parameters d_i , so as to express predator recruitment in terms of net growth rates¹⁰. As such, they give a steady-state growth of zero by definition and so differences in fecundity cannot influence equilibrium coexistence. This does not invalidate the rationale of the Lotka–Volterra model. Our use of separate birth and death terms has revealed the key parameters R_0 and α_{12} that control the ecological impact of male presence on the outcome of competition with asexual mutants. Our conceptual scheme shown in Fig. 1 furthermore indicates that the outcome of competition between populations of consumers may also apply to distributions of species occupying a habitat. If the resource points shown in Fig. 1 are occupied rather than eaten, at a rate r_i , and renew upon local extinction of the colonist at a rate d_i , then our rate equations (4) apply to the well-known Levins model²⁷ of metapopulation dynamics²⁸.

Our application of classical population dynamics has shown how intraspecific competition can halt the invasion of an asexual mutant into a sexual population (Fig. 2). Coexistence is immediately

possible if the invading asexuals have a smaller inhibitory effect on the exploitative abilities of the sexuals than the sexuals have on themselves. The coefficient describing this relative effect, α_{12} , governs a threshold for the growth capacity of the sexual population, below which sexuals cannot compete with the asexual invaders (Fig. 3). Above this threshold, it is self-regulation of asexuals through intraspecific competition that halts their population growth before extinction of the sexual population. We therefore conclude that the twofold cost of males to the growth capacity of the sexual population does not automatically pose a threat to the survival of the sexual strategy. Males are less costly to species with high growth capacities. Strong intraspecific competition favours coexistence after an asexual invasion, giving the sexual strategy more time than was previously thought²⁹ in which to express its well known evolutionary advantages over asex.

Received 9 November 1999; accepted 24 January 2000.

1. Williams, G. C. *Sex and Evolution* (Princeton Univ. Press, Princeton, 1975).
2. Ridley, M. *Evolution* 2nd edn (Blackwell Science, Oxford, 1996).
3. Maynard Smith, J. *The Evolution of Sex* (Cambridge Univ. Press, Cambridge, UK, 1978).
4. Kondrashov, A. S. Classification of hypotheses on the advantage of amphimixis. *J. Hered.* **84**, 372–387 (1993).
5. Judson, O. P. & Normark, B. B. Ancient asexual scandals. *Trends Ecol. Evol.* **11**, 41–46 (1996).
6. Hurst, L. D. & Peck, J. R. Recent advances in understanding of the evolution and maintenance of sex. *Trends Ecol. Evol.* **11**, 46–52 (1996).
7. Hamilton, W. D., Axelrod, R. & Tanese, R. Sexual reproduction as an adaptation to resist parasites. *Proc. Natl Acad. Sci. USA* **87**, 3566–3573 (1990).
8. Volterra, V. in *Animal Ecology* (ed. Chapman, R. N.) 409–448 (McGraw Hill, New York, 1931).
9. Lotka, A. J. The growth of mixed populations: two species competing for a common food supply. *J. Wash. Acad. Sci.* **22**, 461–469 (1932).
10. Begon, M., Harper, J. L. & Townsend, C. R. *Ecology* 3rd edn (Blackwell Science, Oxford, 1996).
11. Vrijenhoek, R. C. Factors affecting clonal diversity and coexistence. *Am. Zool.* **19**, 787–797 (1979).
12. Bell, G. *The Masterpiece of Nature* (Croom Helm, London, 1982).
13. Browne, R. A. & Hoopes, C. W. Genotype diversity and selection in asexual brine shrimp (*Artemia*). *Evolution* **44**, 1035–1051 (1990).
14. Bolger, D. T. & Case, T. J. Divergent ecology of sympatric clones of the asexual gecko, *Lepidodactylus lugubris*. *Oecologia* **100**, 397–405 (1994).

15. Semlitsch, R. D., Hotz, H. & Guex, G. D. Competition among tadpoles of coexisting hemiclones of hybridogenetic *Rana esculenta*: support for the frozen niche variation model. *Evolution* **51**, 1249–1261 (1997).
16. Weeks, A. R. & Hoffmann, A. A. Intense selection of mite clones in a heterogeneous environment. *Evolution* **52**, 1325–1333 (1998).
17. Jokela, J., Lively, C. M., Fox, J. A. & Dybdahl, M. F. Flat reaction norms and “frozen” phenotypic variation in clonal snails (*Potamopyrgus antipodarum*). *Evolution* **51**, 1120–1129 (1997).
18. Vrijenhoek, R. C. & Pfeiler, E. Differential survival of sexual and asexual *Poeciliopsis* during environmental stress. *Evolution* **51**, 1593–1600 (1997).
19. Case, T. J. Patterns of coexistence in sexual and asexual species of *Cnemidophorus* lizards. *Oecologia* **83**, 220–227 (1990).
20. Doebeli, M. An explicit genetic model for ecological character displacement. *Ecology* **77**, 510–520 (1996).
21. Skelton, P. (ed.) *Evolution: a Biological and Palaeontological Approach* (Addison-Wesley, Wokingham, UK, 1993).
22. Bulmer, M. *Theoretical Evolutionary Ecology* (Sinauer, Sunderland, Massachusetts, 1994).
23. Futuyma, D. J. *Evolutionary Biology* 3rd edn (Sinauer, Sunderland, Massachusetts, 1998).
24. Maynard Smith, J. *Evolutionary Genetics* 2nd edn (Oxford Univ. Press, Oxford, 1998).
25. Stearns, S. C. & Hoekstra, R. F. *Evolution: an Introduction* (Oxford Univ. Press, Oxford, 2000).
26. Case, T. J. & Taper, M. L. On the coexistence and coevolution of asexual and sexual competitors. *Evolution* **40**, 366–387 (1986).
27. Levins, R. in *Lectures on Mathematics in the Life Sciences*, vol. 2 (ed. Gerstenhaber, M.) 75–107 (American Mathematical Society, Providence, 1970).
28. Connolly, S. R. & Roughgarden, J. Theory of marine communities: competition, predation, and recruitment-dependent interaction strength. *Ecol. Monogr.* **69**, 277–296 (1999).
29. Lively, C. M. Host-parasite coevolution and sex. *BioScience* **46**, 107–114 (1996).

Acknowledgements

We thank G. F. Turner, S. J. Hawkins and C. P. Please for clarifications of key issues. This work was supported by a Natural Environment Research Council grant to C.P.D., and by an Engineering and Physical Sciences Research Council studentship to G.E.P.

Correspondence should be addressed to C.P.D. (e-mail: cpd@soton.ac.uk).

NMDA spikes in basal dendrites of cortical pyramidal neurons

Jackie Schiller*, Guy Major†‡, Helmut J. Koester‡§ & Yitzhak Schiller‡||

* Departments of Physiology and Biophysics, Technion Medical School, Bat-Galim, Haifa 31096, Israel

† Biological Computation Research-Bell Laboratories, Lucent Technologies, 600 Mountain Avenue, Murray Hill, New Jersey, USA and University Laboratory of Physiology, Oxford, UK

§ Department of Cell Physiology, Max-Planck-Institute for Medical Research, Heidelberg, Germany

|| Department of Neurology, Rambam Medical Center, Haifa, Israel

‡ These authors contributed equally to this work

Basal dendrites are a major target for synaptic inputs innervating cortical pyramidal neurons¹. At present little is known about signal processing in these fine dendrites. Here we show that co-activation of clustered neighbouring basal inputs initiated local dendritic spikes, which resulted in a 5.9 ± 1.5 mV (peak) and 64.4 ± 19.8 ms (half-width) cable-filtered voltage change at the soma that amplified the somatic voltage response by $226 \pm 46\%$. These spikes were accompanied by large calcium transients restricted to the activated dendritic segment. In contrast to conventional sodium or calcium spikes, these spikes were mediated mostly by NMDA (*N*-methyl-D-aspartate) receptor channels, which contributed at least 80% of the total charge. The ionic mechanism of these NMDA spikes may allow ‘dynamic spike-initiation zones’, set by the spatial distribution of glutamate pre-bound to NMDA receptors, which in turn would depend on recent and ongoing activity in the cortical network. In addition, NMDA spikes may serve as a powerful mechanism for modification of the cortical network by inducing long-term strengthening of co-activated neighbouring inputs.

To explore synaptic processing in basal dendrites we identified fine dendritic branches using fluorescence confocal microscopy and mimicked excitatory postsynaptic potentials (EPSPs) by ultraviolet laser glutamate uncaging. Figure 1a compares a synaptically evoked EPSP with an excitatory postsynaptic-like potential (EPSLP) evoked by a 1-ms ultraviolet laser pulse directed to a basal dendrite in the presence of 1 mM extracellular caged glutamate. The two postsynaptic potentials were similar except that EPSLPs were slower and had a larger NMDA component (see Methods for details).

Figure 1b shows EPSLPs evoked in a distal basal dendrite at various laser intensities. At a certain threshold intensity, a small increase in the laser intensity more than doubled the EPSLP amplitude recorded at the soma (Fig. 1b). Beyond that threshold intensity only a small additional increase in EPSLP peak amplitude occurred when the laser power was increased (Fig. 1b, $n = 14$ cells). The threshold and all-or-none nature of the response indicated initiation of local dendritic spikes. The mean amplitude of just subthreshold cable-filtered responses was 3.9 ± 1.1 mV, and the amplitude of the cable-filtered basal dendritic spike was 5.2 ± 1.7 mV, as measured at the soma ($n = 14$). Unless otherwise specified, the terms ‘suprathreshold’ and ‘subthreshold’ are taken to mean ‘suprathreshold in the dendrite’ and ‘subthreshold in the dendrite’, respectively. All EPSLPs presented were subthreshold with respect to axonal action potentials. Hyperpolarization of the membrane potential reduced the amplitudes of the suprathreshold EPSLPs to subthreshold levels (Fig. 1c, $n = 4$). Calcium imaging revealed that suprathreshold EPSLPs evoked a large localized increase in calcium influx limited to a roughly 20- μ m stretch of the activated basal dendrite (Fig. 1d, $n = 5$). These observations are consistent with initiation of a local spike by focal glutamate uncaging onto a single basal dendrite.

Subthreshold EPSLPs were not affected by addition of either the specific voltage-gated sodium channel (VGSC) blocker tetrodotoxin (TTX; 1 μ M) ($96 \pm 10\%$ of control value, $n = 7$) or the voltage-gated calcium channel (VGCC) blocker cadmium (50–150 μ M) ($95 \pm 6\%$ of control value, $n = 10$). In contrast, TTX and cadmium substantially reduced the amplitudes and time integrals of suprathreshold EPSLPs (Fig. 2a, b). The average threshold for basal dendritic spike initiation measured at the soma was 4.2 ± 1.4 mV ($n = 7$) for TTX and 4.4 ± 1.8 mV ($n = 10$) for cadmium. The sodium and calcium-dependent cable-filtered spikes added 5.0 ± 2.1 mV (half-width of 60 ± 16 ms) and 4.7 ± 1.8 mV (half-width of 58 ± 15 ms) to the somatic potential, respectively.

Basal dendritic spikes could be reinitiated following blockade of either VGSCs or VGCCs. The threshold laser intensity for spike initiation increased by $29 \pm 6\%$ in the presence of TTX and $101 \pm 9\%$ in the presence of cadmium. The reinitiated spikes were similar to the original spikes evoked under control conditions (Fig. 2c). The averaged peak and time integral of the reinitiated spikes were $95.7 \pm 8.2\%$ and $117.3 \pm 33.5\%$ of the original spikes ($n = 6$). These results indicated that although VGCCs and VGSCs participated in the initiation of the basal dendritic spike, a third dendritic conductance probably served as the major charge carrier. Furthermore, the fact that basal dendritic spikes could be reinitiated in the presence of either TTX or cadmium ruled out the possibility that these spikes resulted from a polysynaptic or circuit effect.

When glutamate was uncaged at the soma, there were no sudden jumps in response amplitude as laser intensity was increased. Furthermore, TTX and cadmium did not noticeably change somatic EPSLPs ($101 \pm 11\%$ and $104 \pm 4\%$ of control, $n = 6$ and 7, respectively). Thus, somatic VGSCs and VGCCs did not contribute to basal dendritic spikes.

Initiation of local basal dendritic spikes was limited to distal basal dendrites (>70 μ m away from the soma). In three neurons the difference between the distal and proximal dendrite was observed for the same basal dendrite.

Basal dendritic spikes were critically dependent on activation of

NMDA receptors. Figure 3a shows that the addition of the NMDA receptor blocker AP5 (D(-)-2-amino-5-phosphonovaleric acid; 100 μ M) eliminated the basal dendritic spike evoked under control conditions ($n = 12$). Increasing the laser intensity further (up to 775% of the control threshold intensity) did not restore the spike. This was true even when we used high-frequency (40–200 Hz) glutamate uncaging ($n = 5$). In the presence of AP5, the peak and half-width of the AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazolepropionate) receptor-mediated EPSPs were not changed significantly by the addition of cadmium or TTX ($97.8 \pm 5.7\%$ and $105.6 \pm 5.5\%$ of control values, respectively). More importantly, local basal dendritic spikes were initiated even after we blocked both VGSC and VGCC with TTX and cadmium (Fig. 3b). These spikes in turn were blocked by AP5 (Fig. 3c, $n = 5$). 'Pure NMDA spikes' in the presence of TTX and cadmium were higher and wider than the spikes recorded under control conditions (peak amplitude of cable-filtered spikes measured at soma was 7.2 ± 2.6 mV, half-width of 77.3 ± 12.7 ms, $P < 0.01$ compared with control spikes). Our findings indicate that local basal dendritic spikes are mediated for the most part by current flowing through NMDA receptor channels. Sodium and calcium channels assist in spike initiation, but are not necessary for it.

We used computer simulations to quantify the relative contributions of NMDA channels, VGSCs and VGCCs to local basal dendritic spikes (Fig. 3d). With the parameters adjusted so that the model matched the experimental observations, the simulations revealed that NMDA receptor channels contributed around 80% of the charge mediating basal dendritic spikes, with sodium and calcium channels together contributing around 20%. Increasing the density of VGSCs and/or VGCCs so that they contributed more than 20% of the current mediating basal dendritic spikes was inconsistent with the experimental findings that AMPA-mediated

EPSPs were unaffected by cadmium and TTX. The experimental finding that the pure NMDA spike was larger than the control spike was consistent with modelling results showing that initiation of pure NMDA spikes required more NMDA conductance. In addition, the simulations revealed that NMDA spikes involved a 'spike-chain' mechanism: the AMPA response evoked a fast local sodium 'spikelet', which elicited a slower calcium-mediated regenerative response, which in turn initiated the full-blown NMDA spike (Fig. 3d).

To investigate whether local basal dendritic spikes can be evoked by synaptic inputs, we combined focal stimulation of basal dendrites with two-photon imaging. As in the uncaging experiments, when we gradually increased the synaptic stimulus intensity there was an abrupt jump in the EPSP amplitude, and a spike-like response was initiated (Fig. 4a, b). The spike initiation was sensitive to the frequency of activation. Although spikes could be observed with a single EPSP ($n = 6$), they were easier to elicit with high-frequency stimulation (two stimuli at 50–100 Hz; 9 out of 11 neurons examined). At dendritic threshold, the cable-filtered EPSP peak was 2.6 ± 0.8 mV and the cable-filtered basal dendritic spike added 5.9 ± 1.5 mV ($226 \pm 46\%$, half-width 64.4 ± 19.8 ms) to the somatic voltage response ($n = 6$).

The NMDA receptor blocker AP5 (100 μ M) selectively reduced the amplitude of EPSPs that crossed the local dendritic threshold and abolished the spike response (Fig. 4b). Furthermore, AP5 linearized the relationship between the synaptic stimulus intensity and EPSP amplitude and eliminated the abrupt jump observed under control conditions (Fig. 4b, $n = 4$). Basal dendritic spikes could not be reinitiated in the presence of AP5, even with stimulation intensities sufficient to evoke AMPA-mediated EPSPs that were two to three times larger than the just subthreshold EPSPs in control conditions. In control solution sufficient hyperpolarization of the

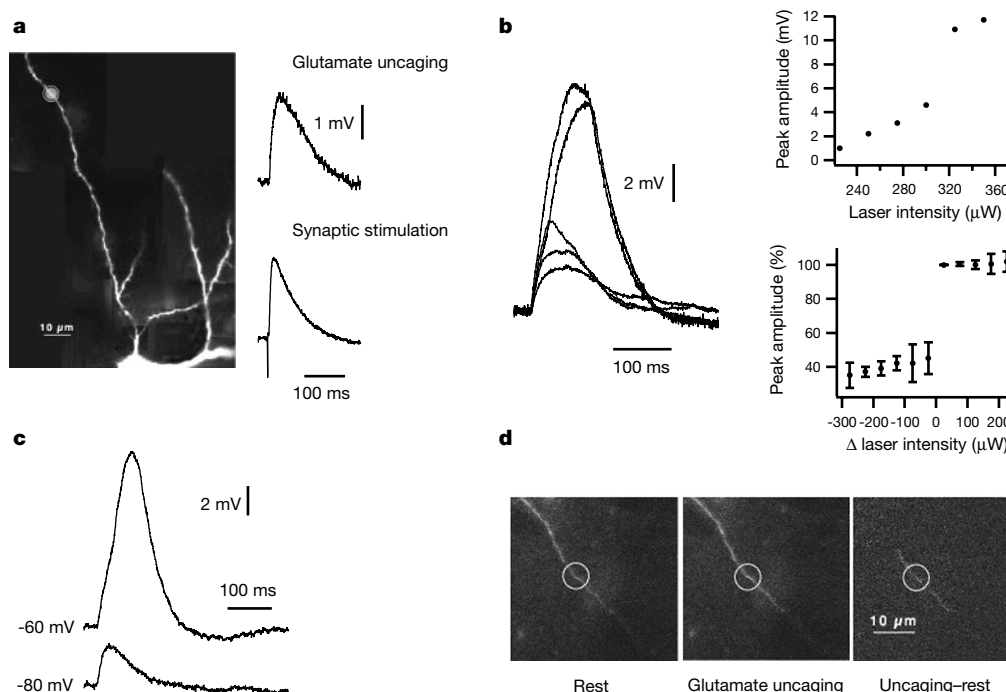


Figure 1 Local spikes in basal dendrites evoked by focal glutamate uncaging. **a**, Fluorescent image of the basal dendritic arborization (left). The grey spot obtained by flash photolysis of caged fluorescein dextran represents the area of glutamate uncaging. Right, somatic voltage responses evoked by glutamate uncaging at a basal dendrite 150 μ m away from the soma (top) and focal synaptic stimulation (bottom) in the same neuron. **b**, Traces of single EPSPs evoked at increasing laser intensities (left). The peak EPSP amplitude is plotted as a function of the laser intensity for the cell shown (top right)

and for an average of 14 cells (bottom right). The laser intensity is shown relative to the laser intensity at threshold (0). The EPSP amplitude is presented in per cent of just supathreshold EPSP (defined as 100%). **c**, The effect of hyperpolarization of the membrane potential on supathreshold EPSP recorded at the soma. **d**, Images of a basal dendritic branch (140 μ m from the soma) at rest and during dendritic spike response. The third image represents the net calcium increase during spike initiation. The circle represents the site of the glutamate uncaging.

membrane potential prevented the dendritic spike response (Fig. 4c, $n = 5$). Together, the last two observations showed that local postsynaptic conductances were responsible for the sharp threshold behaviour, and not recruitment of presynaptic axons or boutons, or polysynaptic circuits.

The similarity between uncaging and synaptically evoked dendritic spikes with respect to amplitude ($P = 0.19$) and half-width ($P = 0.29$), and the fact that synaptically evoked basal dendritic spikes could not be reinitiated in the presence of AP5, indicated that synaptically evoked basal dendritic spikes most probably represent local NMDA spikes. To investigate this further, we performed additional experiments using the intracellular blocker D-890, a predominantly L-type VGCC blocker (Fig. 5a). Initiation of basal dendritic spikes remained intact (peak 5.6 ± 1.14 mV, threshold 2.36 ± 0.38 mV, $n = 4$) in the presence of high concentrations of D-890 (2 mM), despite the fact it reduced the peak of action

potential-evoked calcium transients at basal dendrites by $88 \pm 8\%$ ($n = 4$). These findings indicated further that synaptically evoked basal dendritic spikes represent NMDA rather than calcium spikes.

Simultaneous two-photon calcium imaging from the spines and shafts of the activated site revealed that basal dendritic NMDA spikes were accompanied by a large rise in intradendritic calcium (Fig. 5b). The peak calcium transient versus stimulus intensity curve showed a threshold behaviour with the same threshold value as was found for the voltage responses (compare Figs 4b and 5b). Addition of AP5 significantly reduced the peak calcium transients (by $92 \pm 3\%$ at threshold intensity). Furthermore, as with the voltage responses, AP5 abolished the threshold behaviour, linearizing the intensity response curve (Fig. 5b, $n = 3$). Calcium transients evoked by basal dendritic spikes were restricted to a small segment of a single basal branch (10–20 μm), similar to those described for the

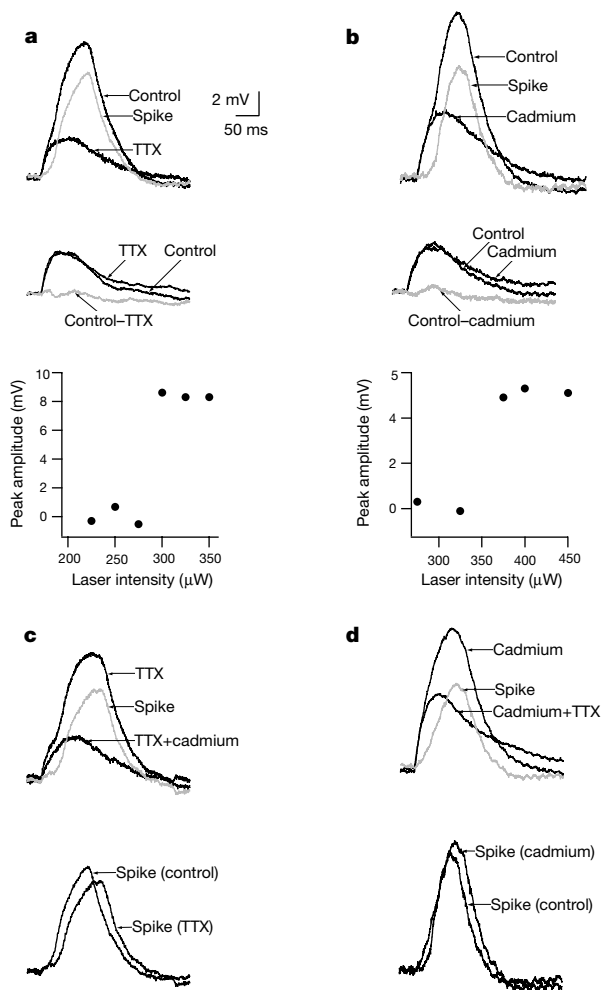


Figure 2 Voltage-gated sodium and calcium channels participate in initiation of basal dendritic spikes. **a, b**, Local subthreshold and suprathreshold EPSPs before (control) and after extracellular application of TTX (1 μM , **a**) and cadmium (100 μM , **b**). The grey traces (control–TTX, control–cadmium, spike) represent a subtraction of traces with the blockers from control traces. Lower panel: the peak of the net sodium- (left, control–TTX) and calcium-dependent (right, control–cadmium) components are plotted as a function of the laser intensities. **c, d**, Reinitiation of spikes in the presence of TTX (1 μM , **c**) and cadmium (100 μM , **d**) upon increasing the laser intensity (29% and 94%, respectively). The reinitiated spikes were blocked in turn by cadmium and TTX, respectively. The traces were obtained from the neurons shown in **a** and **b**, respectively. Lower two traces: The spikes recorded under control conditions are presented for comparison together with the reinitiated spike in the presence of TTX (left) and cadmium (right).

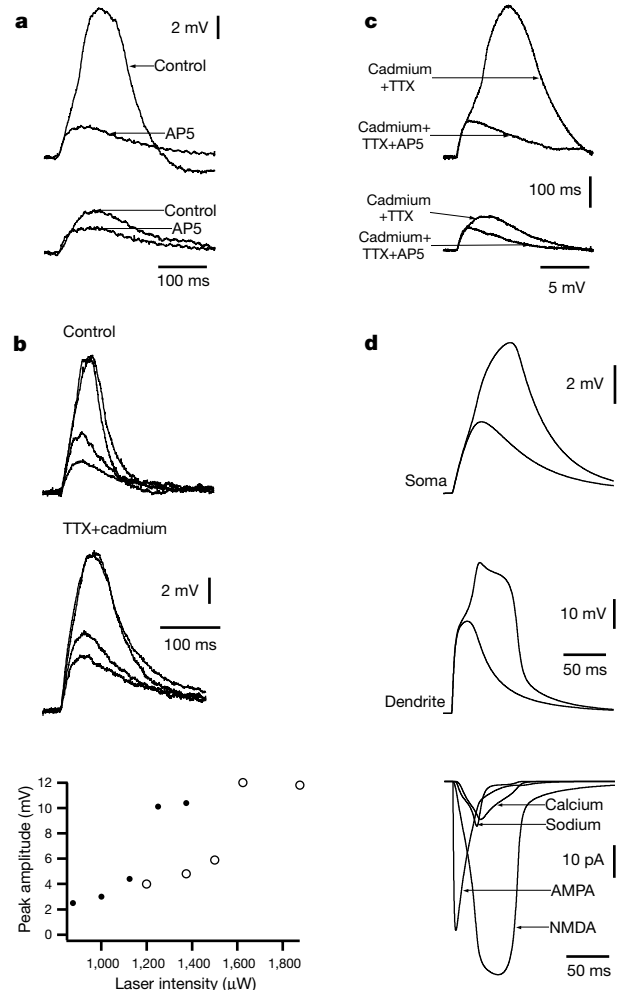


Figure 3 NMDA receptor channels are the main charge carriers mediating basal dendritic spikes. **a**, Subthreshold (lower traces) and suprathreshold (upper traces) EPSPs evoked under control conditions and after addition of 100 μM AP5. **b**, Traces of EPSPs evoked at increasing laser intensities in control conditions and after adding both TTX (1 μM) and cadmium (100 μM , top). Lower panel: peak EPSP amplitudes are plotted as a function of the laser intensity in control conditions (filled circles) and in the presence of the blockers (empty circles). **c**, 'Pure' NMDA spike evoked by glutamate uncaging in the presence of TTX (1 μM) and cadmium (100 μM) before and after the addition of AP5 (100 μM). **d**, Computer simulations of somatic voltage (top), dendritic voltage (middle) and the corresponding dendritic currents (bottom) are given for just subthreshold and just suprathreshold responses. The currents are shown only for the suprathreshold response. The maximum NMDA (nS), sodium and calcium (mS cm^{-2}) conductance densities are G_{NMDA} 11; G_{Na} 2; G_{CaH} 0.02; G_{CaT} 0.15.

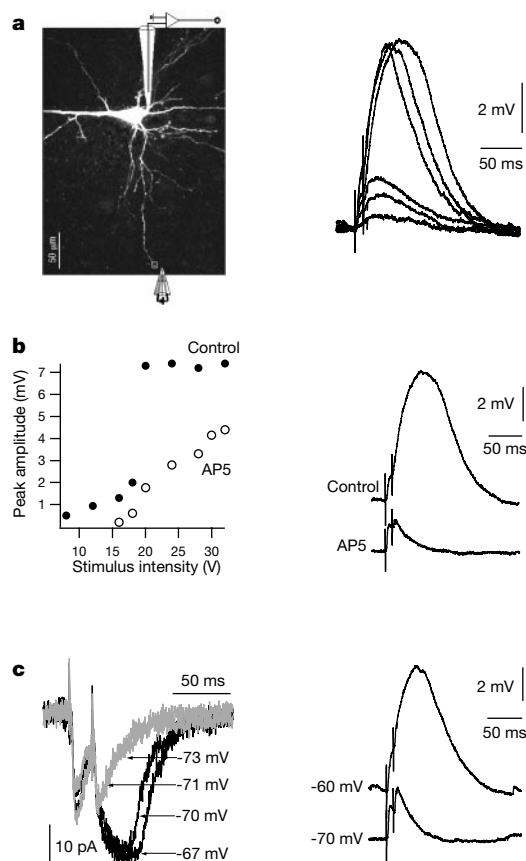


Figure 4 Basal dendritic NMDA spikes evoked by focal synaptic stimulation.

a, Fluorescent image of the basal dendritic arborization. A bipolar theta synaptic stimulating electrode (shown schematically) was placed close to a distal basal dendrite (left). Right, postsynaptic voltage responses (two stimuli at 100 Hz) at different stimulus intensities. **b**, The peaks of postsynaptic voltage responses are plotted as a function of synaptic stimulus intensities under control conditions (filled circles) and in the presence of AP5 (100 μ M, open circles) (left). Basal dendritic spike in control conditions and after addition of AP5 (100 μ M, right). **c**, Synaptically evoked postsynaptic current (voltage-clamp, left) and voltage (current-clamp, right) responses obtained at different hyperpolarizing holding membrane potentials at a single synaptic stimulus intensity that was suprathreshold for spike initiation at the resting membrane potential.

glutamate uncaging experiments. Furthermore, we observed no evidence for activation of neighbouring branches ($n = 3$).

Linear summation of EPSPs within the dendritic trees of CA1 pyramidal neurons has been reported^{2,3}. Although NMDA receptor channels did participate in actively linearizing the summation process, basal dendritic NMDA spikes were not reported. This difference could reflect differences between CA1 and layer 5 cortical neurons, or it could be because, in contrast to the experiments reported here, iontophoretic EPSPs were produced mostly at spatially separated dendritic sites.

NMDA spikes could contribute uniquely to synaptic processing. First, NMDA spikes are localized to the input dendrite, potentially allowing a degree of parallel processing and independent decision making in different basal branches. This is in contrast to the local calcium spikes in the apical tufts of layer 5 neurons, which propagate for several hundred micrometres within the apical dendritic tree⁴. Second, initiation of basal dendritic spikes may be modulated by previous synaptic activity. Pre-bound glutamate on NMDA receptors will, in effect, temporarily turn them into purely voltage-gated channels, and will prime basal dendrites for initiation of NMDA spikes. We speculate that this in turn could allow the cortical circuitry to generate 'dynamic decision zones' on the

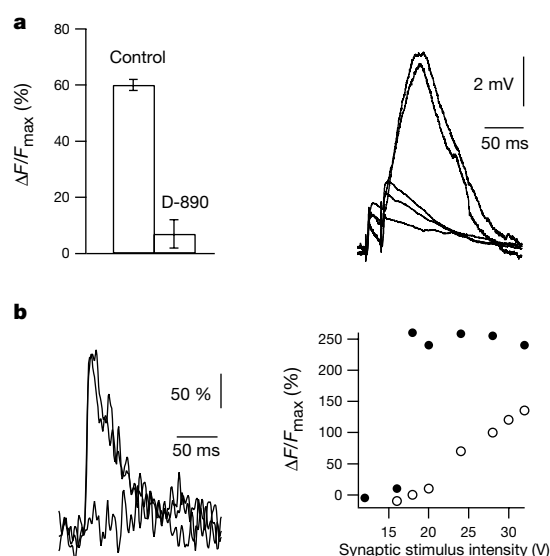


Figure 5 Synaptically evoked basal dendritic spikes are accompanied by large local calcium influx. **a**, Left, averaged peak dendritic calcium transients (presented as $\Delta F/F_{\max}$ in per cent) evoked by somatic action potentials in control conditions and in the presence of D-890 (2 mM) in the pipette solution. Right, postsynaptic voltage responses and dendritic spike initiation evoked by focal basal synaptic stimulation (50 Hz) at different stimulus intensities in the presence of D-890 (2 mM). **b**, Calcium transients obtained from an activated dendritic spine during focal synaptic stimulation at different intensities (left). The corresponding voltage traces are shown in Fig. 4a. Right, the peak calcium transients are presented as a function of the stimulus intensity under control conditions (filled circles) and in the presence of AP5 (open circles).

dendrites of a given neuron. In principle, changes in the ensemble firing pattern of the local neural network could lead to changes in the distribution of primed NMDA conductance 'hot spots' on its constituent neurons. Third, NMDA spikes may provide an efficient mechanism for NMDA-dependent long-term synaptic strengthening⁵ of co-activated neighbouring synapses without the participation of back-propagating action potentials. □

Methods

Slice preparation and electrophysiological recordings

Brain slices 300–350 μ m thick were prepared from sensory-motor areas of 12–35-day-old rat brains as previously described⁶. We carried out whole-cell patch-clamp recordings from visually identified layer 5 pyramidal neurons⁷. The extracellular solution contained (in mM) 125 NaCl, 25 NaHCO₃, 25 glucose, 3 KCl, 1.25 NaH₂PO₄, 2 CaCl₂, 1 MgCl₂, pH 7.4. The intracellular solution contained (in mM) 115 K⁺-gluconate, 20 KCl, 2 Mg-ATP, 2 Na₂-ATP, 10 Na₂-phosphocreatine, 0.3 GTP, 10 HEPES, 0.15–0.5 Calcium Green-1 (CG1) or 0.2 Oregon Green 488 Bapta-1 (OGB-1), pH 7.2. The electrophysiological recordings were performed using an EPC-7 (List-Electronics), and the data were acquired and analysed using PClamp 6.0 (Axon Instruments), Pulse (Heka Elektronik) and Igor (WaveMetrics) software. Focal synaptic stimulation⁸ was performed through a theta patch pipette located close to the selected basal dendritic segment, guided by the fluorescent image of the dendrite. In these experiments the neurons were filled with 200 μ M OGB-1 (Molecular Probes) and the basal dendritic tree was imaged with a confocal imaging system equipped with a two-photon laser (see imaging procedures). In some experiments the calcium channel blocker D-890 (B44280, 2mM, Knoll AG) was added to the recording pipette. Uncaging and synaptic experiments were performed at room temperature and at 36 °C, respectively. The size of the stimulated basal dendritic segment was assessed using calcium imaging and was found to be 10–20 μ m.

Uncaging and imaging procedures

Caged glutamate (γ -ANB-caged L-glutamic acid, Molecular Probes) was photolysed by a 361-nm ultraviolet-laser beam (Innova 300, Coherent) using 1-ms shuttered pulses (UniBlitz shutter driver/timer). After establishing a stable whole-cell recording, we perfused the recording chamber with extracellular solution (bubbled with 95% O₂/5% CO₂) containing freshly prepared caged glutamate compound (1 mM). The basal dendritic arborization was visualized (see imaging techniques for details), the ultraviolet laser was directed to the dendritic region of choice, and brief ultraviolet laser pulses (1 ms) were administered to evoke EPSPs. The full width at half-maximum size of the

photolysed spot was $6.5 \pm 3 \mu\text{m}$ in brain slices, as measured by caged fluorescein dextran (Fig. 1a, dextran BMNB-caged fluorescein, Molecular Probes). Moving the laser excitation horizontally or vertically so that the basal dendrite was more than $15\text{--}20 \mu\text{m}$ from the centre of laser excitation eliminated the EPSP altogether ($n = 5$). Excitatory postsynaptic-like potentials evoked by glutamate uncaging at a rate of $1\text{--}2 \text{ min}^{-1}$ remained stable over the time period tested (up to 40 min, $n = 9$). Blocking both AMPA and NMDA receptors with CNQX ($10 \mu\text{M}$) and AP5 ($50\text{--}150 \mu\text{M}$) eliminated the response during glutamate uncaging ($n = 4$).

Imaging techniques

Neurons were loaded with either CG1 ($150\text{--}500 \mu\text{M}$) or OGB-1 ($200 \mu\text{M}$) through the recording patch pipette and the basal dendritic tree including dendritic spines was imaged. For calcium imaging we used two experimental setups. Experiments using focal glutamate uncaging were performed on an MRC 600 confocal scan head (Bio-Rad) equipped with both an argon ion 488-nm laser (Ion Laser Technology) and a 880-nm two-photon laser (Innova 400 pumping a Mira 900, Coherent) mounted on an upright BX50WI Olympus microscope equipped with a $60\times$ (Olympus 0.9 NA) water objective. Full images were obtained with a temporal resolution of 1 Hz and analysed using Analyse software (Mayo Foundation). Experiments using focal synaptic stimulation were performed using a modified confocal scan head (TCS 4D, Leika Microsystems) mounted on an upright BX50WI Olympus microscope equipped with a $40\times$ (LUMPlanFl 40XW0.9IR, Olympus) water immersion objective. Data were obtained in the line scan mode with a temporal resolution of 4.5 ms and analysed with Igor software⁸.

Modelling procedure

Computer simulations were performed using the compartmental modelling package NEURON⁹ on a Pentium II PC under Windows NT. The modelling was performed on a reconstructed thick tufted layer 5 pyramidal cell. The neuron was reconstructed by A. Larkman as described¹⁰. The cell had 49 basal and 48 apical segments (branches) and a soma diameter of $18.5 \mu\text{m}$. All the results shown are from inputs halfway along the first terminal basal segment. This had a length of $258 \mu\text{m}$ and a diameter of $0.84 \mu\text{m}$ after correction for shrinkage and after the spine collapse procedure. The integration method was CNEXP (Crank-Nicholson Exponential) with a time step of $25 \mu\text{s}$. The temperature (Celsius parameter) was 25°C . Passive parameters were modified from ref. 10 to match the decay time constants and amplitudes of modelled and experimental basal dendritic spikes.

In the model we used fast sodium, intermediate speed low voltage-activated, T-type (CaT) calcium, high voltage-activated calcium (CaH), NMDA and AMPA conductances. A modified Hodgkin-Huxley scheme was used for voltage-sensitive conductances, and the parameters for the voltage gated and synaptic conductances were based on published experimental data^{11–17}. Some of the parameters were modified to allow better fit of the experimental results, and selected model parameters, including conductances (or densities), were varied manually to achieve a good match between the model and the experimental results. NEURON files can be obtained from G.M. (e-mail: gm@physiol.ox.ac.uk).

Received 2 November 1999; accepted 5 January 2000.

1. Larkman, A. U. Dendritic morphology of pyramidal neurons of the rat: Spine distribution. *J. Comp. Neurol.* **306**, 332–343 (1991).
2. Cash, S. & Yuste, R. Input summation by cultured pyramidal neurons is linear and position independent. *J. Neurosci.* **18**, 10–15 (1998).
3. Cash, S. & Yuste, R. Linear summation of excitatory inputs by CA1 pyramidal neurons. *Neuron* **22**, 383–94 (1999).
4. Schiller, J., Schiller, Y., Stuart, G. & Sakmann, B. Calcium action potentials in apical dendrites of neocortical pyramidal neurons in rat brain slices. *J. Physiol.* **505**, 605–616 (1997).
5. Bliss, T. V. P. & Collingridge, G. L. A synaptic model of memory: Long term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
6. Schiller, J., Schiller, Y. & Clapham, D. E. NMDA receptors amplify calcium influx into dendritic spines during associative pre- and postsynaptic activation. *Nature Neurosci.* **1**, 114–118 (1998).
7. Stuart, G. J., Dodt, H. U. & Sakmann, B. Patch-clamp recordings from the soma and dendrites of neurons in brain slices using infrared video microscopy. *Pflug. Arch. Eur. J. Physiol.* **423**, 511–518 (1993).
8. Koester, H. J. & Sakmann, B. Calcium dynamics in single spines during coincident pre- and postsynaptic activity depend on relative timing of back-propagating action potentials and subthreshold excitatory postsynaptic potentials. *Proc. Natl Acad. Sci. USA* **95**, 9596–9601 (1998).
9. Hines, M. L. & Carnevale, N. T. The NEURON simulation environment. *Neural Comput.* **9**, 1179–1209 (1997).
10. Major, G., Larkman, A. U., Jonas, P., Sakmann, B. & Jack, J. J. B. Detailed passive cable models of whole-cell recorded CA3 pyramidal neurons in rat hippocampal slices. *J. Neurosci.* **14**, 4613–4638 (1994).
11. Ascher, P. & Nowak, L. The role of divalent cations in the N-methyl-D-aspartate responses of mouse central neurons in culture. *J. Physiol.* **399**, 247–266 (1988).
12. Hestrin, S., Nicoll, R. A., Perkel, D. J. & Sah, P. Analysis of excitatory synaptic action in pyramidal cells using whole-cell recording from rat hippocampal slices. *J. Physiol.* **422**, 203–225 (1990).
13. Brodin, L., Traven, H. G. C., Lansner, A., Wallen, P. & Ekeberg, O. Computer simulations of N-methyl-D-aspartate receptor-induced membrane properties in a neuron model. *J. Neurophysiol.* **66**, 473–484 (1991).
14. Brown, A. M., Schwindt, P. C. & Crill, W. E. Voltage dependence and activation kinetics of pharmacologically defined components of the high-threshold calcium current in rat neocortical neurons. *J. Neurophysiol.* **70**, 1530–1543 (1993).
15. Jonas, P., Major, G. & Sakmann, B. Quantal components of unitary EPSCs at the mossy fiber synapse on CA3 pyramidal cells of rat hippocampus. *J. Physiol.* **472**, 615–663 (1993).

16. Martina, M. & Jonas, P. Functional differences in Na^+ channel gating between fast-spiking interneurons and principal neurons of rat hippocampus. *J. Physiol.* **505**, 593–603 (1997).
17. Destexhe, A., Neubig, M., Ulrich, D. & Huguenard, J. Dendritic low-threshold calcium currents in thalamic relay cells. *J. Neurosci.* **18**, 3574–3588 (1998).

Acknowledgements

We thank D. E. Clapham, F. Prendergast, B. Sakmann, W. Denk and D. Tank for assistance and support, A. Larkman for reconstructing the cell used in the simulations, and M. Hauser and G. Stuart for reading early versions of the manuscript. In addition we thank the Mayo Foundation, Leo & Frances Kogan endowment fund (J.S.), Wellcome Trust, Lucent Technologies and Marine Biological Laboratory (G.M.) for financial support.

Correspondence and requests for materials should be addressed to J.S. (e-mail: Jackie@tx.technion.ac.il).

The Ras-MAPK pathway is important for olfaction in *Caenorhabditis elegans*

Takaaki Hirotsu, Satoshi Saeki*†, Masayuki Yamamoto* & Yuichi Iino

Molecular Genetics Research Laboratory, The University of Tokyo, Tokyo 113-0033, Japan

* Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan

The Ras-MAPK (mitogen-activated protein kinase) signal transduction pathway is well known to control cellular proliferation and differentiation in response to extracellular signals, but its other functions are less understood. In *Caenorhabditis elegans* this pathway regulates several developmental events, such as vulval induction and progression of meiosis¹, but its function in the nervous system is unknown. Here we report that the Ras-MAPK pathway is involved in olfaction in this organism. Mutational inactivation and hyperactivation of this pathway impairs efficiency of chemotaxis to a set of odorants. Experiments in which *let-60 ras* was expressed using a heat-shock promoter and a cell-specific promoter show that a normal activity of LET-60 Ras is required in mature olfactory neurons. Application of the odorant isoamylalcohol to wild-type animals leads to the activation of MAP kinase in olfactory neurons within 10 seconds. This induction is dependent on the function of the nucleotide-gated channel TAX-2/TAX-4 and the voltage-activated calcium channel subunit UNC-2. These results suggest a dynamic regulatory role for the Ras-MAPK pathway in perception and transmission of sensory signals in olfactory neurons.

Chemotaxis to volatile chemoattractants in *C. elegans* is mediated primarily by two pairs of olfactory neurons, AWA and AWC (ref. 2). The gain-of-function (gf) mutant of *let-60 ras*, *let-60(n1046gf)*, shows a defect in chemotaxis to the AWC-sensed chemoattractants isomylalcohol (Fig. 1a, open bars), butanone and benzaldehyde (data not shown). It also shows a weak but statistically significant defect in chemotaxis to the AWA-sensed chemoattractant diacetyl (Fig. 1b, open bars). This oncogenic-type mutation in *let-60* is expected to lead to a reduced rate of conversion from the GTP-bound active form to the GDP-bound inactive form, and therefore hyperactivation, of the encoded Ras protein³. Defects in chemotaxis to AWA- and AWC-sensed odorants are also seen in the loss-of-function (lf) mutants *let-60(n2021)* and *lin-45(sy96)* and the

† Present address: Tokyo Research Laboratories, Kyowa Hakko Kogyo Co., Ltd, Machidashi, Tokyo 194-8533, Japan.

presumed null mutant *mek-2(n2678)*, where *lin-45* encodes a Raf MAPK kinase kinase and *mek-2* encodes a MAPK kinase, although the defects are very mild in some cases (Fig. 1a, b). These mutants also exhibit defects in chemotaxis to 2,4,5-trimethylthiazole, where AWA and AWC have overlapping functions² (Fig. 1c). It has previously been found that the *mek-2* mutants show poor chemotaxis to *Escherichia coli*⁴, and our results are consistent with this observation. The defect of *let-60(gf)* in chemotaxis to isoamylalcohol is suppressed by loss-of-function mutations in the downstream genes. These include a loss-of-function mutation in the *ksr-1* gene encoding KSR (kinase suppressor of ras)^{5,6} and weak loss-of-function mutations *mek-2(ku114)* and *mpk-1(n2521)*, where *mpk-1* encodes a MAP kinase (Fig. 1d). These results indicate that LET-60 Ras may be involved in chemotaxis to AWA- and AWC-sensed volatile chemoattractants and that the MAPK pathway functions downstream of LET-60 Ras. The defects in these mutants in chemotaxis to odorants are unlikely to be due to a general abnormality in locomotion, because these mutants can respond normally to the water-soluble attractant NaCl, which is sensed mainly by ASE neurons (Fig. 1e).

In vulval induction, SEM-5, an adaptor protein similar to mammalian Grb2, functions upstream of LET-60 Ras⁷, and LIN-1, a transcription factor with partial similarity to mammalian Elk-1, is one of the target molecules of MAPK⁸. The *sem-5(n2019)* and *lin-1(e1777)* mutants show normal chemotaxis to both AWA- and AWC-sensed odorants (data not shown). Thus, these components are unlikely to be shared by the signalling cascade required for

normal chemotaxis.

Because the *let-60(gf)* mutant shows a stronger defect in AWC-mediated chemotaxis, we determined whether there is any abnormality in AWC chemosensory neurons in mutant animals. The *gcy-10* promoter drives expression in the chemosensory neurons AWC and AWB and in the pharyngeal interneuron I1 (ref. 9). We visualized these neurons using the *gcy-10::GFP* marker. A pair of AWC cell bodies was clearly visible in N2 animals transgenic for *gcy-10::GFP*. Another pair of cell bodies, AWB, was also visible, although less bright, whereas I1 appeared very faint. AWC cell bodies were present normally in the *let-60(gf)* and *let-60(lf)* mutants (Fig. 2). No gross morphological abnormalities, such as the generation of new axon branches reported for *tax-2*, *tax-4* and several other mutants¹⁰⁻¹², were noticed in the axons or dendrites.

Given the chemotactic defect in the mutants, an important issue is whether the Ras-MAPK pathway is required for the development or function of differentiated cells. We therefore expressed *let-60* under the control of the heat-shock promoter *hsp16-2*. All the structures essential for chemotaxis differentiate during embryogenesis in *C. elegans* hermaphrodites. We first examined the effect of *let-60(gf)* expression from the *hsp::let-60(gf)* transgene in the wild-type background. Animals in which *let-60(gf)* was expressed during embryogenesis showed a normal response to isoamylalcohol and diacetyl (Fig. 3a). In contrast, chemotaxis to isoamylalcohol and to diacetyl decreased significantly when *let-60(gf)* was expressed in adults (Fig. 3b, c). These results indicate that the ectopic activation of Ras in fully differentiated cells may impair chemotaxis. Again, chemotaxis to NaCl was not affected in these animals (data not shown).

In reciprocal experiments, the expression of *let-60(+)* as a transgene in the *let-60(lf)* mutant during adulthood restored the response to isoamylalcohol and diacetyl significantly (Fig. 3d, e). Although the rescue was not complete, possibly because of unstable maintenance of the transgene as an extrachromosomal array or because of suboptimal expression levels, this result is consistent with chemotaxis requiring the normal function of *let-60* in differentiated cells.

To determine the site of action of *let-60*, we tested whether the expression of *let-60(gf)* in AWC chemosensory neurons is sufficient

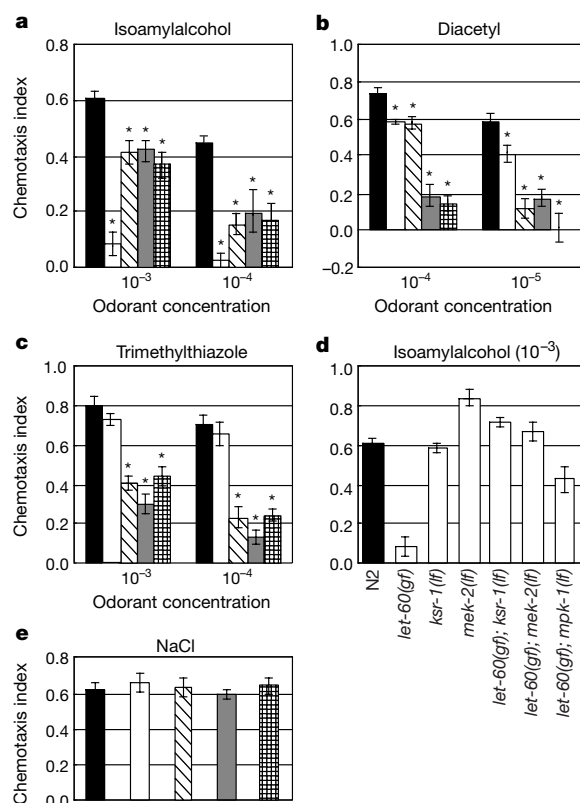


Figure 1 Chemotaxis of mutants affected in the Ras-MAPK pathway. **a**, Chemotaxis to isoamylalcohol, which is sensed mainly by AWC chemosensory neurons, of wild-type N2 (filled bars), *let-60(n1046gf)* (open bars), *let-60(n2021lf)* (hatched bars), *lin-45(sy96lf)* (stippled bars) and *mek-2(n2678lf)* (cross-hatched bars). The attractant dilutions used are indicated at the bottom. **b**, Chemotaxis to diacetyl, which is sensed mainly by AWA neurons, depicted as in **a**. **c**, Chemotaxis to 2,4,5-trimethylthiazole, sensed by AWA and AWC neurons, depicted as in **a**. **d**, Chemotaxis to isoamylalcohol at 10^{-3} dilution. Strains: wild-type N2, *let-60(n1046)*, *ksr-1(ku68)*, *mek-2(ku114)*, *let-60(n1046); ksr-1(ku68)*, *mek-2(ku114); let-60(n1046)* and *mpk-1(n2521); let-60(n1046)*. **e**, Chemotaxis to 100 mM sodium chloride. Bars are labelled as in **a**.

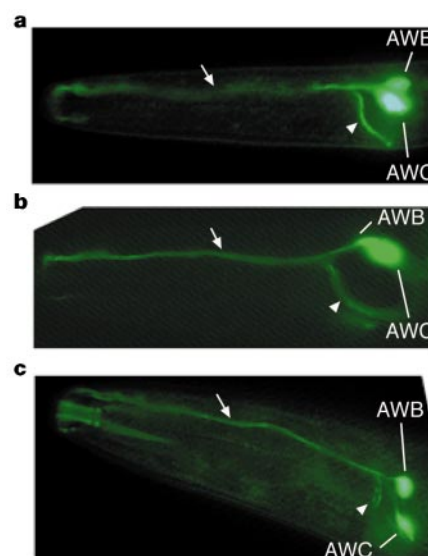


Figure 2 AWC neurons visualized with the *gcy-10::GFP* marker. Green fluorescent protein was seen in the AWC and AWB cell bodies (as indicated) in N2 (wild type) (**a**), *let-60(n1046gf)* mutant (**b**) and *let-60(n2021lf)* mutant (**c**). The dendrites (arrows) and axons (arrowheads) of these neurons were also visible and showed no gross morphological abnormalities. The anterior tip of the animal, where sensory cilia are located, is to the left in each panel.

to cause the defect in chemotaxis to AWC-sensed odorants. We placed *let-60(gf)* under the control of the *gcy-10* promoter, which acts in AWC, AWB and I1 neurons, and introduced it into wild-type animals. The transgenic lines showed a significant defect in chemotaxis to isoamylalcohol, which is sensed by AWC, but not to diacetyl, which is sensed by AWA (Fig. 3f). Similarly, *gcy-10::let-60(+)* restored the response of the *let-60(lf)* mutant to isoamylalcohol, but not to diacetyl (Fig. 3g). These results indicate that AWC neurons may be responsible for the chemotactic defects in both *let-60(gf)* and *let-60(lf)* mutants. Note that AWB chemosensory neurons, which also express *gcy-10*, sense volatile repellents and are dispensable for chemotaxis to volatile attractants^{2,13}, and that pharyngeal neurons, including I1, are mostly isolated from the somatic nervous system¹⁴.

Our genetic analysis shows that the activity of the Ras-MAPK pathway in mature olfactory neurons is important for chemotaxis to AWC-sensed odorants. How, then, would the Ras-MAPK pathway be activated in olfactory neurons? To assess the activity of the Ras-MAPK pathway directly, we used a monoclonal antibody to activated (diphosphorylated) MAP kinase. AWC neurons were identi-

fied by co-staining with anti-GFP antibodies in the *gcy-10::GFP* transformants. Immunoreactivity against the anti-activated-MAPK antibody was observed in the AWC neurons, along with several other neurons including AWB, after 10 seconds of application of isoamylalcohol at 10^{-2} or 10^{-4} dilution (Fig. 4a, b). The most intense staining is observed in the cell bodies of these neurons. Axons and dendrites also show weak but consistent staining, whereas no staining was ever observed in the sensory cilia. Only a few animals showed staining in the AWC neurons without odorant stimulus (Fig. 4b). In the *let-60(gf)* mutant, many neurons in the head region including AWC show intense staining irrespective of odorant application. In contrast, no staining was observed in the *let-60(lf)* mutants even after odorant stimulus (Fig. 4a, b). These observations indicate that the immunoreactivity detected with this antibody is a true MAP kinase activated downstream of LET-60 Ras.

We further examined the involvement of possible upstream components in this response. In AWC neurons, olfactory signals received by seven-transmembrane receptors are thought to be transmitted through G protein to the cGMP pathway. Mutations in the *odr-3* gene encoding a G protein α -subunit and in the *tax-2* or

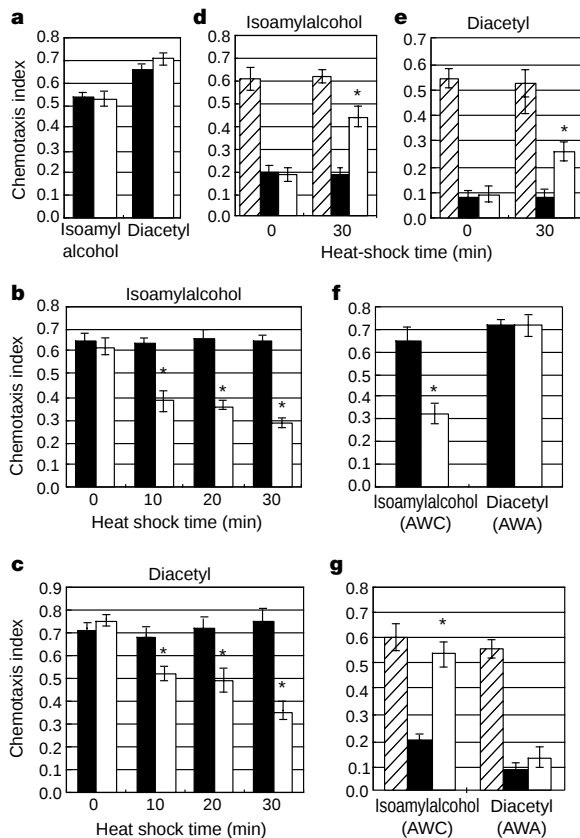


Figure 3 Temporally and spatially restricted expression of *let-60* affects chemotaxis to odorants. **a–c**, Heat shock was applied during embryogenesis (**a**) or adulthood (**b, c**) to wild-type animals transgenic for *hsp::let-60(gf)*, and chemotaxis to the indicated odorants was assayed (open bars). Control animals carried only pEF1::GFP, the transformation marker (filled bars). **d–e**, The *let-60(n2021lf)* mutant carrying *hsp::let-60(+)* (open bars) or only pEF1::GFP (filled bars) were subjected to heat shock as adults, and chemotaxis to isoamylalcohol (**d**) and diacetyl (**e**) was assayed. Hatched bars show wild-type animals carrying only pEF1::GFP. **f**, Wild-type animals carrying *gcy-10::let-60(gf)* (open bars) and those carrying only pEF1::GFP (filled bars) were assayed for chemotaxis to isoamylalcohol and diacetyl. **g**, The *let-60(n2021lf)* mutant carrying *gcy-10::let-60(+)* (open bars) and that carrying only pEF1::GFP (filled bars) were assayed for chemotaxis to isoamylalcohol and diacetyl. Hatched bars show wild-type animals carrying only pEF1::GFP. The dilution of isoamylalcohol was 10^{-3} (**a, b, f**) and 10^{-4} (**d, g**). The dilution of diacetyl was 10^{-4} (**a, c, f**) and 10^{-5} (**e, g**).

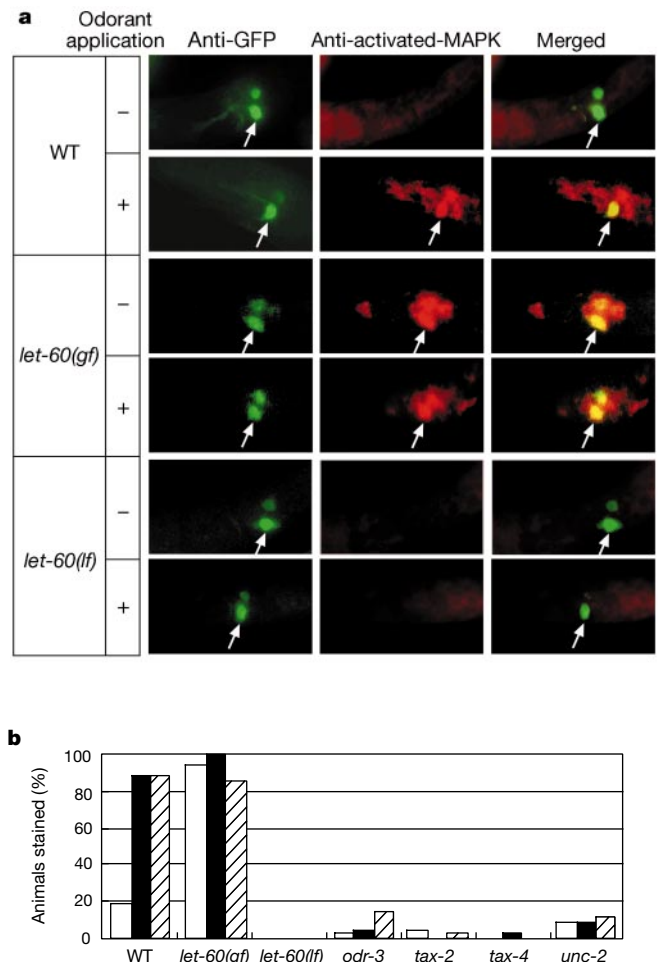


Figure 4 Activation of MAPK by odorant stimulus. **a**, Immunofluorescence images of anti-GFP staining (green), anti-activated-MAPK staining (red) and merged images (yellow indicates overlap). All figures show the left lateral views of the head regions of animals. '-', animals fixed without odorant application; '+', animals fixed 10 s after application of 10^{-2} dilution of isoamylalcohol. Arrows show AWC neurons. **b**, The proportion of animals showing staining with anti-activated-MAPK in AWC neurons. Animals were exposed for 10 s to isoamylalcohol at a concentration of 10^{-2} (filled bars) or 10^{-4} (hatched bars). Open bars show control experiments without application of the odorant. Strains: wild-type N2, *let-60(n1046gf)*, *let-60(n2021lf)*, *odr-3(n2150lf)*, *tax-2(p691lf)*, *tax-4(p678lf)*, *unc-2(e55lf)*. All were transgenic for *gcy-10::GFP*.

tax-4 genes encoding subunits of the cyclic nucleotide-gated cation channel cause defects in AWC-mediated chemotaxis^{11,15,16}. Activation of MAP kinase by the application of isoamylalcohol is severely reduced or eliminated in any of these mutants (Fig. 4b). *unc-2* encodes a protein similar to the α_1 -subunits of neuronal voltage-activated calcium channels¹⁷. Activation of the MAP kinase is also severely reduced in the *unc-2* mutant. These results suggest that odorant-induced neuronal activity is essential for the activation of the Ras-MAPK pathway.

This study provides the first evidence that the Ras-MAPK pathway functions in the nervous system of *C. elegans*. It is also the first observation, to our knowledge, that indicates a functional role for the Ras-MAPK pathway in olfactory neurons in any species. In *C. elegans*, as in vertebrates¹⁸, olfactory signals are thought to be received by seven-transmembrane receptors and transmitted to the cGMP pathway, and to lead to cation influx through the TAX-2/TAX-4 channel in AWC olfactory neurons¹⁹. Our study has shown that the Ras-MAPK pathway functions downstream of these components.

Unexpectedly, MAPK was activated after only ten seconds of odorant stimulus in the chemosensory neuron cell bodies, which are nearly 100 μm away from the sensory cilia. This fast response, together with our observation that sensory cilia are devoid of activated MAPK, indicates that electrical activity may mediate the response. There is good evidence that *C. elegans* lack voltage-activated sodium channels, and that a depolarization-activated calcium current is essential for the maintenance of membrane depolarization^{20,21}. The loss of the odorant-induced activation of MAPK in the *unc-2* mutant is, therefore, probably due to a failure to sustain electrical activity in the sensory neurons. In addition, calcium ions may be directly involved in the activation, possibly through a calcium-activated guanine nucleotide exchange factor similar to mammalian Ras-GRF (refs 22–24).

Our results show that both hyperactivation and inactivation of the pathway impairs chemotaxis. This suggests that Ras and the downstream MAPK should be properly activated in response to odorants and shut off in their absence, and that failure to do so leads to dysfunction of the sensory neurons. The pathway may, therefore, be directly involved in some aspect of synaptic transmission, such as regulation of synaptic vesicle supply. Otherwise it may have a regulatory role in chemosensation, such as adjustment of the dynamic range of the sensory system depending on the odorant concentration. The effects of MAPK could be through either transcriptional regulation or the direct phosphorylation of target molecules. Given the power of genetics in *C. elegans*, suppressor or enhancer screens should reveal downstream genes encoding targets or downstream effectors of MAPK. □

Methods

Chemotaxis assay

Chemotaxis assays for volatile odorants were performed on phosphate-buffered agar plates as described², except in the assay for diacetyl, for which MOPS-buffered agar plates were used²⁵. Washed animals were placed ~3 cm from the point where the odorants had been spotted. Chemotaxis to NaCl was assayed essentially as described²⁶ on a phosphate-buffered assay plate in which a chemical gradient was formed for 15 h by an agar plug including 100 mM NaCl. Roughly 100 adult hermaphrodites grown at 20 °C were tested at 20 °C on each assay plate. All the strains used were viable and fertile as homozygotes and maintained as such, except for *mek-2(n2678)*, which were sterile as homozygotes. The *mek-2* allele was maintained as heterozygotes and only sterile progeny were counted in the chemotaxis assay. In Figs 1 and 3, each bar represents at least five independent assays. Error bars show s.e.m. and asterisks indicate that the values differ significantly ($P < 0.05$) from that of the control. Chemotaxis index = ((number of animals attracted to the chemoattractant) – (number of animals at the opposite end of the plate))/(total number of animals on the assay plate).

Germline transformation

Plasmids were injected into the gonad arms as described²⁷. For visualization of the AWC neurons, the *gcy-10::GFP* construct (from D. Garbers) was injected with the plasmid pRF4, which contains a dominant allele of *rol-6*, as a transformation marker.

Heat shock experiments

Complementary DNAs corresponding to the wild-type *let-60* allele and to the *let-60(n1046gf)* allele (both from M. Han) were inserted into pPD49.78 (from A. Fire) which carries the heat-shock promoter *hsp16-2*. The resulting *hsp::let-60(gf)* was injected into N2 (wild type), and *hsp::let-60(+)* was injected into *let-60(n2021lf)* at a concentration of 50 $\mu\text{g ml}^{-1}$ using pEF1 α ::GFP (from N. Hisamoto; 50 $\mu\text{g ml}^{-1}$) as a fluorescent transformation marker. This transformation marker did not affect chemotaxis or locomotion. For embryonic expression of the transgene, gravid hermaphrodites were collected and treated with alkaline hypochlorite to separate the embryos physically. Early embryos thus isolated were heat shocked at 30 °C for 8 h, and allowed to grow to adulthood at 20 °C, after which chemotaxis assay was conducted. For expression in adults, 10–30 min of heat shock at 33 °C was applied to adults suspended in buffer, and the chemotaxis assay was performed after recovery at 20 °C for 1 h. Only animals emitting green fluorescence under a fluorescence-equipped dissecting microscope (Olympus SZX12), which indicates the maintenance of the unstable extrachromosomal array, were counted in these assays.

Cell-specific expression

The promoter region (1.7 kb) of *gcy-10* was amplified by PCR from *C. elegans* genomic DNA and inserted into the pPD49.26 vector into which *let-60(+)* or *let-60(gf)* cDNA had been inserted. The *gcy-10::let-60* constructs were injected with pEF1 α ::GFP as a transformation marker at a concentration of 50 $\mu\text{g ml}^{-1}$ each.

Immunohistochemistry

Animals were collected and kept in water for 3 h before odorant application. Excess water was removed after centrifugation and 50 μl of isoamylalcohol diluted in water was added. After 10 s, we added Bowin's fixative/Methanol/BME (40:40:1) and carried out immunofluorescence²⁸ using a fluorescent microscope (Zeiss Axioskop). Incubation with primary antibodies was at room temperature for 15 h, and secondary antibodies were incubated at room temperature for 2 h. The primary antibodies used were mouse anti-diphosphorylated-ERK1&2 monoclonal antibody (Sigma M8159)²⁹ at 1/800, and rabbit anti-GFP polyclonal antibodies (Clontech) at 1/800. The secondary antibodies were Cy3-conjugated goat anti-mouse IgG antibodies (Jackson ImmunoResearch) at 1/200, and FITC-conjugated goat anti-rabbit IgG (Sigma) at 1/200. In Fig. 4b, each bar represents a total of two independent experiments (number of animals scored; 10–20) for *let-60(gf)* and *let-60(lf)*, and three independent experiments ($n = 27$ –67) for other strains.

Received 6 December 1999; accepted 20 January 2000.

1. Sternberg, P. W. & Han, M. Genetics of RAS signaling in *C. elegans*. *Trends Genet.* **14**, 466–472 (1998).
2. Bargmann, C. I., Hartwig, E. & Horvitz, H. R. Odorant-selective genes and neurons mediate olfaction in *C. elegans*. *Cell* **74**, 515–527 (1993).
3. Beitel, G. J., Clark, S. G. & Horvitz, H. R. *Caenorhabditis elegans* *ras* gene *let-60* acts as a switch in the pathway of vulval induction. *Nature* **348**, 503–509 (1990).
4. Church, D. L., Guan, K. L. & Lambie, E. J. Three genes of the MAP kinase cascade, *mek-2*, *mpk-1/sur-1* and *let-60 ras*, are required for meiotic cell cycle progression in *Caenorhabditis elegans*. *Development* **121**, 2525–2535 (1995).
5. Kornfeld, K., Hom, D. B. & Horvitz, H. R. The *ksr-1* gene encodes a novel protein kinase involved in Ras-mediated signaling in *C. elegans*. *Cell* **83**, 903–913 (1995).
6. Sundaram, M. & Han, M. The *C. elegans ksr-1* gene encodes a novel Raf-related kinase involved in Ras-mediated signal transduction. *Cell* **83**, 889–901 (1995).
7. Clark, S. G., Stern, M. J. & Horvitz, H. R. *C. elegans* cell-signalling gene *sem-5* encodes a protein with SH2 and SH3 domains. *Nature* **356**, 340–344 (1992).
8. Tan, P. B., Lackner, M. R. & Kim, S. K. MAP kinase signaling specificity mediated by the LIN-1 Ets/LIN-31 WH transcription factor complex during *C. elegans* vulval induction. *Cell* **93**, 569–580 (1998).
9. Yu, S., Avery, L., Baude, E. & Garbers, D. L. Guanylyl cyclase expression in specific sensory neurons: A new family of chemosensory receptors. *Proc. Natl Acad. Sci. USA* **94**, 3384–3387 (1997).
10. Coburn, C. M., Mori, I., Ohshima, Y. & Bargmann, C. I. A cyclic nucleotide-gated channel inhibits sensory axon outgrowth in larval and adult *Caenorhabditis elegans*: a distinct pathway for maintenance of sensory axon structure. *Development* **125**, 249–258 (1998).
11. Coburn, C. M. & Bargmann, C. I. A putative cyclic nucleotide-gated channel is required for sensory development and function in *C. elegans*. *Neuron* **17**, 695–706 (1996).
12. Peckol, E. L., Zellen, J. A., Yarrow, J. C. & Bargmann, C. I. Sensory activity affects sensory axon development in *C. elegans*. *Development* **126**, 1891–1902 (1999).
13. Troemel, E. R., Kimmel, B. E. & Bargmann, C. I. Reprogramming chemotaxis responses: sensory neurons define olfactory preferences in *C. elegans*. *Cell* **91**, 161–169 (1997).
14. Albertson, D. G. & Thomson, J. N. The pharynx of *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **275**, 299–325 (1976).
15. Roayaie, K., Crump, J. G., Sagasti, A. & Bargmann, C. I. The G alpha protein ODR-3 mediates olfactory and nociceptive function and controls cilium morphogenesis in *C. elegans* olfactory neurons. *Neuron* **20**, 55–67 (1998).
16. Komatsu, H., Mori, I., Rhee, J. S., Akaike, N. & Ohshima, Y. Mutations in a cyclic nucleotide-gated channel lead to abnormal thermosensation and chemosensation in *C. elegans*. *Neuron* **17**, 707–718 (1996).
17. Schafer, W. R. & Kenyon, C. J. A calcium-channel homologue required for adaptation to dopamine and serotonin in *Caenorhabditis elegans*. *Nature* **375**, 73–78 (1995).
18. Schild, D. & Restrepo, D. Transduction mechanisms in vertebrate olfactory receptor cells. *Physiol. Rev.* **78**, 429–466 (1998).
19. Prasad, B. C. & Reed, R. R. Chemosensation: molecular mechanisms in worms and mammals. *Trends Genet.* **15**, 150–153 (1999).
20. Bargmann, C. I. Neurobiology of the *Caenorhabditis elegans* genome. *Science* **282**, 2028–2033 (1998).
21. Goodman, M. B., Hall, D. H., Avery, L. & Lockery, S. R. Active currents regulate sensitivity and dynamic range in *C. elegans* neurons. *Neuron* **20**, 763–772 (1998).
22. Finkbeiner, S. & Greenberg, M. E. Ca^{2+} -dependent routes to Ras: mechanisms for neuronal survival, differentiation, and plasticity? *Neuron* **16**, 233–236 (1996).

23. Farnsworth, C. L. *et al.* Calcium activation of Ras mediated by neuronal exchange factor Ras-GRE. *Nature* **376**, 524–527 (1995).
24. Brambilla, R. *et al.* A role for the Ras signalling pathway in synaptic transmission and long-term memory. *Nature* **390**, 281–286 (1997).
25. Dusenbery, D. B., Sheridan, R. E. & Russell, R. L. Chemotaxis-defective mutants of the nematode *Caenorhabditis elegans*. *Genetics* **80**, 297–309 (1975).
26. Bargmann, C. I. & Horvitz, H. R. Chemosensory neurons with overlapping functions direct chemotaxis to multiple chemicals in *C. elegans*. *Neuron* **7**, 729–742 (1991).
27. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extra-chromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970 (1991).
28. Nonet, M. L. *et al.* *Caenorhabditis elegans* *rab-3* mutant synapses exhibit impaired function and are partially depleted of vesicles. *J. Neurosci.* **17**, 8061–8073 (1997).
29. Yung, Y. *et al.* Detection of ERK activation by a novel monoclonal antibody. *FEBS Lett.* **408**, 292–296 (1997).

Acknowledgements

We thank D. Garbers for the *gcy-10::GFP* reporter plasmid; M. Han for *let-60* cDNAs and the *mek-2(ku114)* strain; M. Koga, Y. Ohshima, N. Hisamoto and K. Matsumoto for pEF1α::GFP; A. Fire for vectors; and C. Bargmann, T. Schedl and Y. Emori for their comments and advice. All other nematode strains used in this study were provided by the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources (NCRR).

Correspondence and requests for materials should be addressed to Y.I. (e-mail: iino@ims.u-tokyo.ac.jp).

An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells

Scott M. Hammond*, Emily Bernstein†‡, David Beach*§ & Gregory J. Hannon‡

* *Genetica, Inc., P.O. Box 99, Cold Spring Harbor, New York 11724, USA*

† *Graduate Program in Genetics, State University of New York at Stony Brook, Stony Brook, New York 11794, USA*

§ *Wolfson Institute for Biological Sciences, University College London, Gower Street, London WC1E 6BT, UK*

‡ *Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA*

In a diverse group of organisms that includes *Caenorhabditis elegans*, *Drosophila*, planaria, hydra, trypanosomes, fungi and plants, the introduction of double-stranded RNAs inhibits gene expression in a sequence-specific manner^{1–7}. These responses, called RNA interference or post-transcriptional gene silencing, may provide anti-viral defence, modulate transposition or regulate gene expression^{1,6,8–10}. We have taken a biochemical approach towards elucidating the mechanisms underlying this genetic phenomenon. Here we show that 'loss-of-function' phenotypes can be created in cultured *Drosophila* cells by transfection with specific double-stranded RNAs. This coincides with a marked reduction in the level of cognate cellular messenger RNAs. Extracts of transfected cells contain a nuclease activity that specifically degrades exogenous transcripts homologous to transfected double-stranded RNA. This enzyme contains an essential RNA component. After partial purification, the sequence-specific nuclease co-fractionates with a discrete, ~25-nucleotide RNA species which may confer specificity to the enzyme through homology to the substrate mRNAs.

Although double-stranded RNAs (dsRNAs) can provoke gene silencing in numerous biological contexts including *Drosophila*^{11,12}, the mechanisms underlying this phenomenon have remained mostly unknown. We therefore wanted to establish a biochemically tractable model in which such mechanisms could be investigated.

Transient transfection of cultured, *Drosophila* S2 cells with a *lacZ* expression vector resulted in β-galactosidase activity that was easily

detectable by an *in situ* assay (Fig. 1a). This activity was greatly reduced by co-transfection with a dsRNA corresponding to the first 300 nucleotides of the *lacZ* sequence, whereas co-transfection with a control dsRNA (*CD8*) (Fig. 1a) or with single-stranded RNAs of either sense or antisense orientation (data not shown) had little or no effect. This indicated that dsRNAs could interfere, in a sequence-specific fashion, with gene expression in cultured cells.

To determine whether RNA interference (RNAi) could be used to target endogenous genes, we transfected S2 cells with a dsRNA corresponding to the first 540 nucleotides of *Drosophila cyclin E*, a gene that is essential for progression into S phase of the cell cycle. During log-phase growth, untreated S2 cells reside primarily in G2/M (Fig. 1b). Transfection with *lacZ* dsRNA had no effect on cell-cycle distribution, but transfection with the *cyclin E* dsRNA caused a G1-phase cell-cycle arrest (Fig. 1b). The ability of *cyclin E* dsRNA to provoke this response was length-dependent. Double-stranded RNAs of 540 and 400 nucleotides were quite effective, whereas

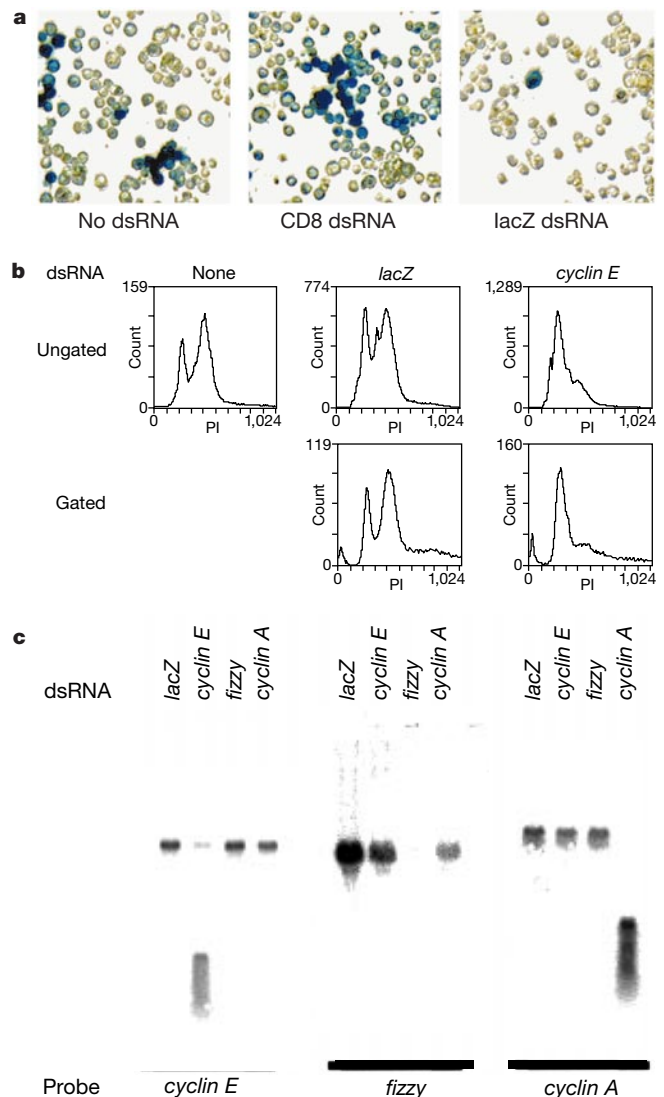


Figure 1 RNAi in S2 cells. **a**, *Drosophila* S2 cells were transfected with a plasmid that directs *lacZ* expression from the copia promoter in combination with dsRNAs corresponding to either human CD8 or *lacZ*, or with no dsRNA, as indicated. **b**, S2 cells were co-transfected with a plasmid that directs expression of a GFP-US9 fusion protein (12) and dsRNAs of either *lacZ* or *cyclin E*, as indicated. Upper panels show FACS profiles of the bulk population. Lower panels show FACS profiles from GFP-positive cells. **c**, Total RNA was extracted from cells transfected with *lacZ*, *cyclin E*, *fizzy* or *cyclin A* dsRNAs, as indicated. Northern blots were hybridized with sequences not present in the transfected dsRNAs.

dsRNAs of 200 and 300 nucleotides were less potent. Double-stranded *cyclin E* RNAs of 50 or 100 nucleotides were inert in our assay, and transfection with a single-stranded, antisense *cyclin E* RNA had virtually no effect (see Supplementary Information).

One hallmark of RNAi is a reduction in the level of mRNAs that are homologous to the dsRNA. Cells transfected with the *cyclin E* dsRNA (bulk population) showed diminished endogenous *cyclin E* mRNA as compared with control cells (Fig. 1c). Similarly, transfection of cells with dsRNAs homologous to *fizzy*, a component of the anaphase-promoting complex (APC) or *cyclin A*, a cyclin that acts in S, G2 and M, also caused reduction of their cognate mRNAs (Fig. 1c). The modest reduction in *fizzy* mRNA levels in cells transfected with *cyclin A* dsRNA probably resulted from arrest at a point in the division cycle at which *fizzy* transcription is low^{14,15}. These results indicate that RNAi may be a generally applicable method for probing gene function in cultured *Drosophila* cells.

The decrease in mRNA levels observed upon transfection of specific dsRNAs into *Drosophila* cells could be explained by effects at transcriptional or post-transcriptional levels. Data from other systems have indicated that some elements of the dsRNA response may affect mRNA directly (reviewed in refs 1 and 6). We therefore sought to develop a cell-free assay that reflected, at least in part, RNAi.

S2 cells were transfected with dsRNAs corresponding to either *cyclin E* or *lacZ*. Cellular extracts were incubated with synthetic mRNAs of *lacZ* or *cyclin E*. Extracts prepared from cells transfected with the 540-nucleotide *cyclin E* dsRNA efficiently degraded the *cyclin E* transcript; however, the *lacZ* transcript was stable in these lysates (Fig. 2a). Conversely, lysates from cells transfected with the *lacZ* dsRNA degraded the *lacZ* transcript but left the *cyclin E* mRNA intact. These results indicate that RNAi ablates target mRNAs through the generation of a sequence-specific nuclease activity. We have termed this enzyme RISC (RNA-induced silencing complex).

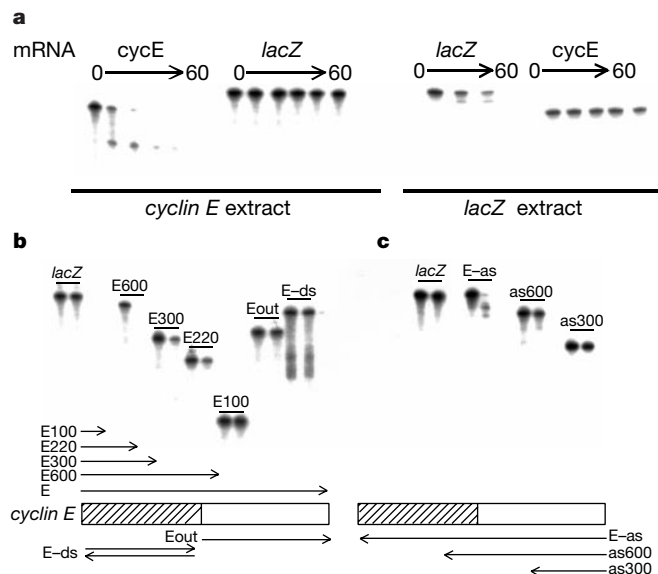


Figure 2 RNAi *in vitro*. **a**, Transcripts corresponding to either the first 600 nucleotides of *Drosophila cyclin E* (E600) or the first 800 nucleotides of *lacZ* (Z800) were incubated in lysates derived from cells that had been transfected with either *lacZ* or *cyclin E* (*cycE*) dsRNAs, as indicated. Time points were 0, 10, 20, 30, 40 and 60 min for *cyclin E* and 0, 10, 20, 30 and 60 min for *lacZ*. **b**, Transcripts were incubated in an extract of S2 cells that had been transfected with *cyclin E* dsRNA (cross-hatched box, below). Transcripts corresponded to the first 800 nucleotides of *lacZ* or the first 600, 300, 220 or 100 nucleotides of *cyclin E*, as indicated. Eout is a transcript derived from the portion of the *cyclin E* cDNA not contained within the transfected dsRNA. E-ds is identical to the dsRNA that had been transfected into S2 cells. Time points were 0 and 30 min. **c**, Synthetic transcripts complementary to the complete *cyclin E* cDNA (Eas) or the final 600 nucleotides (Eas600) or 300 nucleotides (Eas300) were incubated in extract for 0 or 30 min.

Although we occasionally observed possible intermediates in the degradation process (see Fig. 2), the absence of stable cleavage end-products indicates an exonuclease (perhaps coupled to an endonuclease). However, it is possible that the RNAi nuclease makes an initial endonucleolytic cut and that non-specific exonucleases in the extract complete the degradation process¹⁶. In addition, our ability to create an extract that targets *lacZ* *in vitro* indicates that the presence of an endogenous gene is not required for the RNAi response.

To examine the substrate requirements for the dsRNA-induced, sequence-specific nuclease activity, we incubated a variety of *cyclin E*-derived transcripts with an extract derived from cells that had been transfected with the 540-nucleotide *cyclin E* dsRNA (Fig. 2b, c). Just as a length requirement was observed for the transfected dsRNA, the RNAi nuclease activity showed a dependence on the size of the RNA substrate. Both a 600-nucleotide transcript that extends slightly beyond the targeted region (Fig. 2b) and an ~1-kilobase (kb) transcript that contains the entire coding sequence (data not shown) were completely destroyed by the extract. Surprisingly, shorter substrates were not degraded as efficiently. Reduced activity was observed against either a 300- or a 220-nucleotide transcript, and a 100-nucleotide transcript was resistant to nuclease in our assay. This was not due solely to position effects because ~100-nucleotide transcripts derived from other portions of the transfected dsRNA behaved similarly (data not shown). As expected, the nuclease activity (or activities) present in the extract could also recognize the antisense strand of the *cyclin E* mRNA. Again, substrates that contained a substantial portion of the targeted region were degraded efficiently whereas those that contained a shorter stretch of homologous sequence (~130 nucleotides) were recognized inefficiently (Fig. 2c, as600). For both the sense and antisense strands, transcripts that had no homology with the transfected dsRNA (Fig. 2b, Eout; Fig. 2c, as300) were not degraded. Although we cannot exclude the possibility that nuclease specificity could have migrated beyond the targeted region, the resistance of transcripts that do not contain homology to the dsRNA is consistent with data from *C. elegans*. Double-stranded RNAs homologous to an upstream cistron have little or no effect on a linked downstream cistron, despite the fact that unprocessed, polycistronic mRNAs can be readily detected^{17,18}. Furthermore, the nuclease was inactive against a dsRNA identical to that used to provoke the RNAi response *in vivo* (Fig. 2b). In the *in vitro* system, neither a 5' cap nor a poly(A) tail was required, as such transcripts were degraded as efficiently as uncapped and non-polyadenylated RNAs.

Gene silencing provoked by dsRNA is sequence specific. A plausible mechanism for determining specificity would be incorporation of nucleic-acid guide sequences into the complexes that accomplish silencing¹⁹. In accord with this idea, pre-treatment of

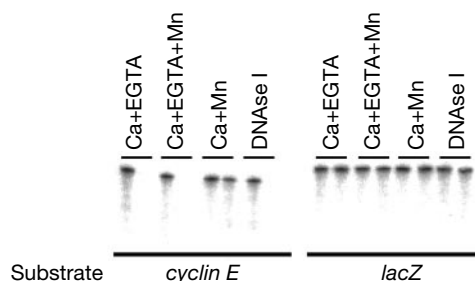


Figure 3 Substrate requirements of the RISC. Extracts were prepared from cells transfected with *cyclin E* dsRNA. Aliquots were incubated for 30 min at 30 °C before the addition of either the *cyclin E* (E600) or *lacZ* (Z800) substrate. Individual 20- μ l aliquots, as indicated, were pre-incubated with 1 mM CaCl_2 and 5 mM EGTA, 1 mM CaCl_2 , 5 mM EGTA and 60 U of micrococcal nuclease, 1 mM CaCl_2 and 60 U of micrococcal nuclease or 10 U of DNase I (Promega) and 5 mM EGTA. After the 30-min pre-incubation, EGTA was added to those samples that lacked it. Yeast tRNA (1 μ g) was added to all samples. Time points were at 0 and 30 min.

extracts with a Ca^{2+} -dependent nuclease (micrococcal nuclease) abolished the ability of these extracts to degrade cognate mRNAs (Fig. 3). Activity could not be rescued by addition of non-specific RNAs such as yeast transfer RNA. Although micrococcal nuclease can degrade both DNA and RNA, treatment of the extract with DNase I had no effect (Fig. 3). Sequence-specific nuclease activity, however, did require protein (data not shown). Together, our results support the possibility that the RNAi nuclease is a ribonucleoprotein, requiring both RNA and protein components. Biochemical fractionation (see below) is consistent with these components being associated in extract rather than being assembled on the target mRNA after its addition.

In plants, the phenomenon of co-suppression has been associated with the existence of small (~ 25 -nucleotide) RNAs that correspond to the gene that is being silenced¹⁹. To address the possibility that a similar RNA might exist in *Drosophila* and guide the sequence-specific nuclease in the choice of substrate, we partially purified our activity through several fractionation steps. Crude extracts contained both sequence-specific nuclease activity and abundant, heterogeneous RNAs homologous to the transfected dsRNA (Figs 2 and 4a). The RNAi nuclease fractionated with ribosomes in a high-speed centrifugation step. Activity could be extracted by treatment with high salt, and ribosomes could be removed by an additional centrifugation step. Chromatography of soluble nuclease over an anion-exchange column resulted in a discrete peak of activity (Fig. 4b, *cyclin E*). This retained specificity as it was inactive against a heterologous mRNA (Fig. 4b, *lacZ*). Active fractions also contained an RNA species of 25 nucleotides that is homologous to the *cyclin E* target (Fig. 4b, northern). The band observed on northern blots may represent a family of discrete RNAs because it could be detected with probes specific for both the sense and antisense *cyclin E* sequences and with probes derived from distinct segments of the dsRNA (data not shown). At present, we cannot determine whether the 25-nucleotide RNA is present in the nuclease complex in a double-stranded or single-stranded form.

RNA interference allows an adaptive defence against both exogenous and endogenous dsRNAs, providing something akin to a dsRNA immune response. Our data, and that of others¹⁹, is con-

sistent with a model in which dsRNAs present in a cell are converted, either through processing or replication, into small specificity determinants of discrete size in a manner analogous to antigen processing. Our results suggest that the post-transcriptional component of dsRNA-dependent gene silencing is accomplished by a sequence-specific nuclease that incorporates these small RNAs as guides that target specific messages based upon sequence recognition. The identical size of putative specificity determinants in plants¹⁹ and animals predicts a conservation of both the mechanisms and the components of dsRNA-induced, post-transcriptional gene silencing in diverse organisms. In plants, dsRNAs provoke not only post-transcriptional gene silencing but also chromatin remodelling and transcriptional repression^{20,21}. It is now critical to determine whether conservation of gene-silencing mechanisms also exists at the transcriptional level and whether chromatin remodelling can be directed in a sequence-specific fashion by these same dsRNA-derived guide sequences. □

Note added in proof: Recently, Tuschl *et al.* have reported the development of cell-free extracts from *Drosophila* embryos that can carry out RNAi (T. Tuschl, P. D. Zamore, D. P. Bartel and P. A. Sharp, *Genes Dev.* **13**, 3191–3197; 1999). Their results also indicate that the RNAi is accomplished at least in part by nuclease degradation of targeted mRNAs.

Methods

Cell culture and RNA methods

S2 (ref. 22) cells were cultured at 27 °C in 90% Schneider's insect media (Sigma), 10% heat inactivated fetal bovine serum (FBS). Cells were transfected with dsRNA and plasmid DNA by calcium phosphate co-precipitation²³. Identical results were observed when cells were transfected using lipid reagents (for example, Superfect, Qiagen). For FACS analysis, cells were additionally transfected with a vector that directs expression of a green fluorescent protein (GFP)–US9 fusion protein¹³. These cells were fixed in 90% ice-cold ethanol and stained with propidium iodide at 25 $\mu\text{g ml}^{-1}$. FACS was performed on an Elite flow cytometer (Coulter). For northern blotting, equal loading was ensured by over-probing blots with a control complementary DNA (RP49). For the production of dsRNA, transcription templates were generated by polymerase chain reaction such that they contained T7 promoter sequences on each end of the template. RNA was prepared using the RiboMax kit (Promega). Confirmation that RNAs were double stranded came from their complete sensitivity to RNase III (a gift from A. Nicholson). Target mRNA transcripts were synthesized using the Riboprobe kit (Promega) and were gel purified before use.

Extract preparation

Log-phase S2 cells were plated on 15-cm tissue culture dishes and transfected with 30 μg dsRNA and 30 μg carrier plasmid DNA. Seventy-two hours after transfection, cells were harvested in PBS containing 5 mM EGTA washed twice in PBS and once in hypotonic buffer (10 mM HEPES pH 7.3, 6 mM β -mercaptoethanol). Cells were suspended in 0.7 packed-cell volumes of hypotonic buffer containing Complete protease inhibitors (Boehringer) and 0.5 units ml^{-1} of RNasin (Promega). Cells were disrupted in a dounce homogenizer with a type B pestle, and lysates were centrifuged at 30,000g for 20 min. Supernatants were used in an *in vitro* assay containing 20 mM HEPES pH 7.3, 110 mM KOAc, 1 mM $\text{Mg}(\text{OAc})_2$, 3 mM EGTA, 2 mM CaCl_2 , 1 mM DTT. Typically, 5 μl extract was used in a 10 μl assay that contained also 10,000 c.p.m. synthetic mRNA substrate.

Extract fractionation

Extracts were centrifuged at 200,000g for 3 h and the resulting pellet (containing ribosomes) was extracted in hypotonic buffer containing also 1 mM MgCl_2 and 300 mM KOAc. The extracted material was spun at 100,000g for 1 h and the resulting supernatant was fractionated on Source 15Q column (Pharmacia) using a KCl gradient in buffer A (20 mM HEPES pH 7.0, 1 mM dithiothreitol, 1 mM MgCl_2). Fractions were assayed for nuclease activity as described above. For northern blotting, fractions were proteinase K/SDS treated, phenol extracted, and resolved on 15% acrylamide 8M urea gels. RNA was electroblotted onto Hybond N+ and probed with strand-specific riboprobes derived from *cyclin E* mRNA. Hybridization was carried out in 500 mM NaPO_4 pH 7.0, 15% formamide, 7% SDS, 1% BSA. Blots were washed in 1 $\times$ SSC at 37–45 °C.

Received 26 November 1999; accepted 26 January 2000.

1. Sharp, P. A. RNAi and double-strand RNA. *Genes Dev.* **13**, 139–141 (1999).
2. Sanchez-Alvarado, A. & Newmark, P. A. Double-stranded RNA specifically disrupts gene expression during planarian regeneration. *Proc. Natl Acad. Sci. USA* **96**, 5049–5054 (1999).
3. Lohmann, J. U., Endl, I. & Bosch, T. C. Silencing of developmental genes in Hydra. *Dev. Biol.* **214**, 211–214 (1999).
4. Cogoni, C. & Macino, G. Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature* **399**, 166–169 (1999).
5. Waterhouse, P. M., Graham, M. W. & Wang, M. B. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc. Natl Acad. Sci. USA* **95**, 13959–13964 (1998).

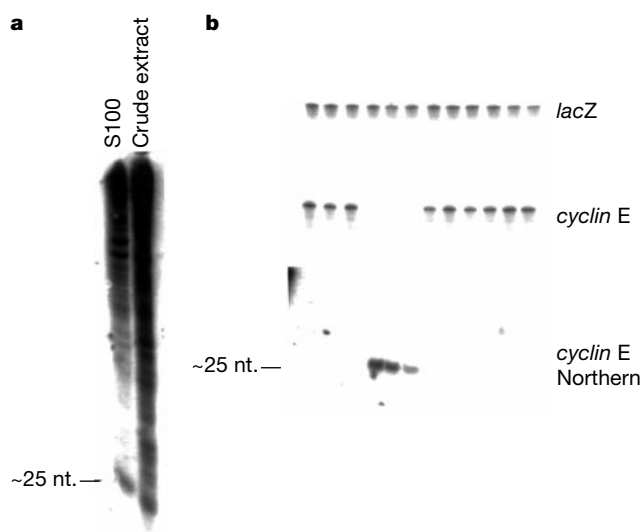


Figure 4 The RISC contains a potential guide RNA. **a**, Northern blots of RNA from either a crude lysate or the S100 fraction (containing the soluble nuclease activity, see Methods) were hybridized to a riboprobe derived from the sense strand of the *cyclin E* mRNA. **b**, Soluble *cyclin-E*-specific nuclease activity was fractionated as described in Methods. Fractions from the anion-exchange resin were incubated with the *lacZ*, control substrate (upper panel) or the *cyclin E* substrate (centre panel). Lower panel, RNA from each fraction was analysed by northern blotting with a uniformly labelled transcript derived from sense strand of the *cyclin E* cDNA. DNA oligonucleotides were used as size markers.

6. Montgomery, M. K. & Fire, A. Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression. *Trends Genet.* **14**, 225–228 (1998).
7. Ngo, H., Tschudi, C., Gull, K. & Ullu, E. Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc. Natl Acad. Sci. USA* **95**, 14687–14692 (1998).
8. Tabara, H. *et al.* The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**, 123–132 (1999).
9. Ketting, R. F., Haverkamp, T. H. A., van Luenen, H. G. A. M. & Plasterk, R. H. A. *mut-7* of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner Syndrome helicase and RnaseD. *Cell* **99**, 133–141 (1999).
10. Ratcliff, F., Harrison, B. D. & Baulcombe, D. C. A similarity between viral defense and gene silencing in plants. *Science* **276**, 1558–1560 (1997).
11. Kennerdell, J. R. & Carthew, R. W. Use of dsRNA-mediated genetic interference to demonstrate that *frizzled* and *frizzled 2* act in the wingless pathway. *Cell* **95**, 1017–1026 (1998).
12. Misquitta, L. & Paterson, B. M. Targeted disruption of gene function in *Drosophila* by RNA interference: a role for nautilus in embryonic somatic muscle formation. *Proc. Natl Acad. Sci. USA* **96**, 1451–1456 (1999).
13. Kalejta, R. F., Brideau, A. D., Banfield, B. W. & Beavis, A. J. An integral membrane green fluorescent protein marker, U59-GFP, is quantitatively retained in cells during propidium iodide-based cell cycle analysis by flow cytometry. *Exp. Cell. Res.* **248**, 322–328 (1999).
14. Wolf, D. A. & Jackson, P. K. Cell cycle: oiling the gears of anaphase. *Curr. Biol.* **8**, R637–R639 (1998).
15. Kramer, E. R., Gieffers, C., Holz, G., Hengstschlager, M. & Peters, J. M. Activation of the human anaphase-promoting complex by proteins of the CDC20/fizzy family. *Curr. Biol.* **8**, 1207–1210 (1998).
16. Shuttlesworth, J. & Colman, A. Antisense oligonucleotide-directed cleavage of mRNA in *Xenopus* oocytes and eggs. *EMBO J.* **7**, 427–434 (1988).
17. Tabara, H., Grishok, A. & Mello, C. C. RNAi in *C. elegans*: soaking in the genome sequence. *Science* **282**, 430–432 (1998).
18. Boshier, J. M., Dufourcq, P., Sookharea, S. & Labouesse, M. RNA interference can target pre-mRNA. Consequences for gene expression in a *Caenorhabditis elegans* operon. *Genetics* **153**, 1245–1256 (1999).
19. Hamilton, J. A. & Baulcombe, D. C. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**, 950–952 (1999).
20. Jones, L. A., Thomas, C. L. & Maule, A. J. *De novo* methylation and co-suppression induced by a cytoplasmically replicating plant RNA virus. *EMBO J.* **17**, 6385–6393 (1998).
21. Jones, L. A. *et al.* RNA–DNA interactions and DNA methylation in post-transcriptional gene silencing. *Plant Cell* **11**, 2291–2301 (1999).
22. Schneider, I. Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *J. Embryol. Exp. Morphol.* **27**, 353–365 (1972).
23. Di Nocera, P. P. & Dawid, I. B. Transient expression of genes introduced into cultured cells of *Drosophila*. *Proc. Natl Acad. Sci. USA* **80**, 7095–7098 (1983).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank C. Velinon and L. Rodgers for assistance with flow cytometry. Materials and advice were provided by A. Krainer, J. Yin and A. Nicholson. D.B. is supported by the Hugh and Catherine Stevenson Fund. G.J.H. is a Pew Scholar in the Biomedical Sciences. This work was supported in part by grants from the NIH (G.J.H.) and the US Army Breast Cancer Research Program (G.J.H.).

Correspondence and requests for materials should be addressed to G.J.H. (e-mail: hannon@cshl.org).

A genetic link between co-suppression and RNA interference in *C. elegans*

René F. Ketting* & Ronald H. A. Plasterk*

Division of Molecular Biology, The Netherlands Cancer Institute, Centre for Biomedical Genetics, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Originally discovered in plants^{1,2}, the phenomenon of co-suppression by transgenic DNA has since been observed in many organisms from fungi³ to animals^{4–7}: introduction of transgenic copies of a gene results in reduced expression of the transgene as well as the endogenous gene. The effect depends on sequence identity between transgene and endogenous gene. Some cases of co-suppression resemble RNA interference (the experimental silencing of genes by the introduction of double-stranded RNA)⁸,

as RNA seems to be both an important initiator and a target in these processes^{9–13}. Here we show that co-suppression in *Caenorhabditis elegans* is also probably mediated by RNA molecules. Both RNA interference^{14,15} and co-suppression¹⁶ have been implicated in the silencing of transposons. We now report that mutants of *C. elegans* that are defective in transposon silencing and RNA interference (*mut-2*, *mut-7*, *mut-8* and *mut-9*) are in addition resistant to co-suppression. This indicates that RNA interference and co-suppression in *C. elegans* may be mediated at least in part by the same molecular machinery, possibly through RNA-guided degradation of messenger RNA molecules.

We tested whether the MUT-7 protein, a putative 3'–5' exoribonuclease required for transposon silencing and RNA interference (RNAi)¹⁴, is also required for co-suppression in *C. elegans*. Co-suppression in *C. elegans* has been reported for a number of genes, including *fem-1*. As described previously⁷, wild-type animals bearing a highly repetitive transgene containing multiple copies of the complete *fem-1* gene show a feminization of the germline, phenocopying loss-of-function mutations of the *fem-1* gene (Table 1). It has been shown that this effect depends on the presence of the *fem-1* promoter region⁷. When this region is not present, no feminization is observed, indicating that RNA is a mediator in co-suppression. We placed the same *fem-1* transgene in a *mut-7* mutant background and found that this feminization was no longer observed (Table 1). This result indicates that the RNA-mediated co-suppression effect of the *fem-1* transgene has a genetic basis and that it requires a protein (MUT-7) that is also involved in the processes of RNAi and transposon silencing. Thus, the aberrant RNA molecules that have been postulated in co-suppression^{10,17} might be double-stranded RNA (dsRNA) molecules, also involved in RNAi⁸.

To test whether the dependence of co-suppression on *mut-7* is general, we analysed two other genes for which co-suppression effects have been described: *gld-1* (ref. 6) and *mrt-2* (S. Ahmed and J. Hodgkin, personal communication). *Gld-1* co-suppression leads to an absence of oocytes and a tumorous germline, whereas *mrt-2* co-suppression results in hypersensitivity to ionizing radiation (which is consistent with the loss-of-function phenotypes of both genes^{18,19}). Again, we find that the observed co-suppression effects

Table 1 Co-suppression of *fem-1*

Genotype	No. of animals with phenotype	
	Feminized	Wild type
Wild type; pKEx1534	28	2
<i>mut-7(pk204)</i> ; pKEx1534	0	27
<i>rde-1(ne219)</i> ; pKEx1539	30	1

Feminization of the germline by transgenes containing the *fem-1* gene. pKEx1534 was generated by injection of *fem-1* plasmid DNA into *mut7(pk204)* animals. This resulted in several non-co-suppressed transgenic lines (one containing the transgene pKEx1534). Restoration of *mut-7* gene function results in feminization of the germline. Injection of the same DNA into *rde-1(ne219)* animals results in lines displaying high levels of feminization.

Table 2 Co-suppression of *gld-1*

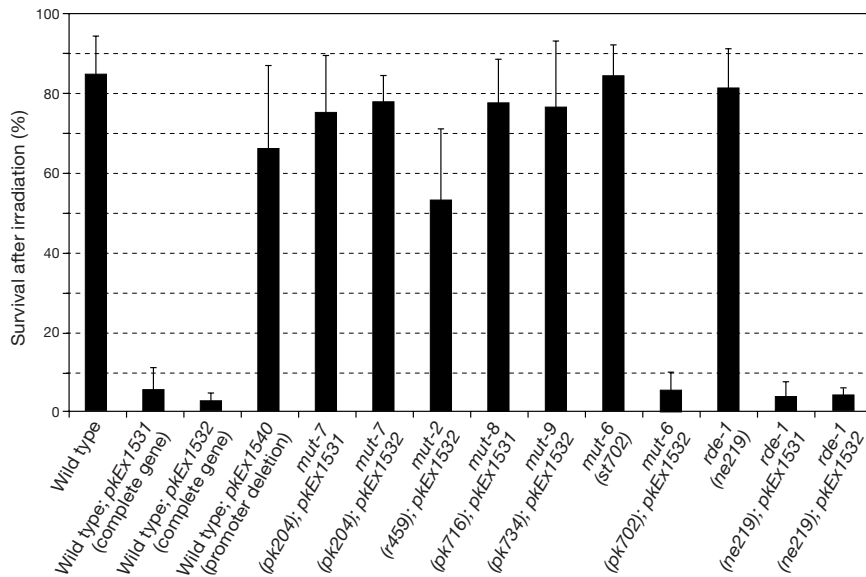
Genotype	No. of animals with phenotype	
	Tumorous germline	Wild type
<i>mut-7(pk204)</i> ; pKEx1533	32	4
<i>mut-7(pk204)</i> ; pKEx1533	4	35
Complete promoter in wild type*	7	0
Complete promoter in <i>rde-1(ne219)</i> *	4	0
Deleted promoter in wild type*	0	3†
Promoter only in wild type*	0	11

Induction of a 'tumorous germline' phenotype¹⁸ by a *gld-1* multicopy transgene (pKEx1533), containing the complete *gld-1* promoter.

*The number of stable non-co-suppressed lines (designated wild type) or co-suppressed lines (tumorous germline) after injection is given; 31 (complete promoter in wild type), 22 (complete promoter in *rde-1(ne219)*), 32 (deleted promoter) and 107 (promoter only) F₁ transgenic animals were analysed.

†These three lines have no tumorous germline and produce oocytes. The strains produce some unfertilized eggs, indicative of a sperm defect, probably caused by a lower dosage of GLD-1 protein¹⁸.

* Present address: Hubrecht Laboratory, Centre for Biomedical Genetics, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.



mut-2, *mut-8* and *mut-9* abolishes the *mrt-2* co-suppression effect. These mutants are both mutator and resistant to RNAi. Mutator activity (*mut-6(st702)*) or RNAi resistance (*rde-1(ne219)*) alone is not sufficient to interfere with *mrt-2* co-suppression.

thus presumably acts via RNAi, is affected by the *rde-1(ne219)* mutation¹⁵. These results indicate that co-suppression is not the same as RNA interference.

It has been recognized previously that co-suppression effects in plants and *Drosophila* can involve transcriptional gene silencing (TGS) or post-transcriptional gene silencing (PTGS) mechanisms. PTGS has been correlated with RNA breakdown^{22,23}, whereas TGS involves DNA-related processes such as methylation^{24,25} (although it is also found in PTGS^{26,27}) and binding of polycomb-group proteins to co-suppressed genes^{4,5}. Our results show co-suppression phenotypes for only those mutants that affect RNAi and transposon silencing together; these mutants seem to affect DNA as well as RNA (the Him phenotype involves chromosome segregation, the Rde phenotype RNA breakdown). This observation supports models in which co-suppression acts via DNA as well as RNA^{12,28,29}. Furthermore, our results are in agreement with our previously proposed model of transposon detection through dsRNA¹⁴, because transcription of transgenes seems to be a prerequisite for co-suppression. It is too early to tie all observations together into one model; molecular identification and further characterization of other *rde* and *mut* genes should make this possible. Assuming that the genes analysed have a direct mechanistic role (as proposed for *mut-7* (ref. 14)), one could speculate that the activation of an RNA-degrading complex through dsRNA directs chromatin changes to chromosomal locations with homology to the dsRNA, to prevent the further production of repetitive mRNA molecules. Whatever the mechanisms, it is an attractive hypothesis that RNAi and co-suppression have evolved to silence repetitive sequences that might invade the genome of an organism. □

Methods

Transgenic animals were obtained through microinjection as described³⁰. The total DNA concentration was 80 ng μl^{-1} ; 50 ng μl^{-1} co-suppressing plasmid (CB#1095 for *fem-1* (ref. 7), pAJ37 for *gld-1* (ref. 6), pRP1240 for promoter deletion of *gld-1*, pY41C4.14 for *mrt-2* (ref. 19), pRP1243 for promoter deletion of *mrt-2* and pRP1244 for *gld-1* promoter) and 30 ng μl^{-1} pRF4 (ref. 30). Multiple independent transgenes were generated in each case, all of which showed phenotypes comparable to those presented here. pRP1240 was derived from pAJ37 by deletion of a 1-kilobase (kb) *Hind*III fragment. This left 69 base pairs intact before the ATG. This 1-kb *Hind*III fragment was cloned into the *Hind*III site of pGEM3 to give pRP1243. pRP1243 was derived from pY41C4.14 by deleting a 3-kb *Xho*I fragment. Extrachromosomal arrays were made in NL917(*mut-7(pk204)III*), except where

NATURE | VOL 404 | 16 MARCH 2000 | www.nature.com

indicated. The *mrt-2* arrays were crossed into the various genetic backgrounds via Bristol N2.

To determine the sensitivity of the *mrt-2* transgenic strains to ionizing radiation, L4 larvae were irradiated with 60 Gy, using a ¹³⁷Cs source. After 48 h, single animals were selected; their progeny was scored for survival 24 h later.

Received 28 October 1999; accepted 16 February 2000.

1. Napoli, C., Lemieux, C. & Jorgensen, R. A. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell* **2**, 279–289 (1990).
2. van der Krol, A. R., Mur, L. A., Beld, M., Mol, J. N. & Stuitje, A. R. Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell* **2**, 291–299 (1990).
3. Cogoni, C. & Macino, G. Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature* **399**, 166–169 (1999).
4. Pal-Bhadra, M., Bhadra, U. & Birchler, J. A. Cosuppression in *Drosophila*: gene silencing of alcohol dehydrogenase by white-Adh transgenes is polycycom dependent. *Cell* **90**, 385–387 (1997).
5. Pal-Bhadra, M., Bhadra, U. & Birchler, J. A. Cosuppression of nonhomologous transgenes in *Drosophila* involves mutually related endogenous sequences. *Cell* **99**, 35–46 (1999).
6. Jones, A. R. & Schedl, T. Mutations in *gld-1*, a female germ cell-specific tumor suppressor gene in *Caenorhabditis elegans*, affect a conserved domain also found in Src-associated protein Sam68. *Genes Dev.* **9**, 1491–1504 (1995).
7. Gaudet, J., van der Elst, I. & Spence, A. M. Post-transcriptional regulation of sex determination in *Caenorhabditis elegans*: widespread expression of the sex-determining gene *fem-1* in both sexes. *Mol. Biol. Cell* **7**, 1107–1121 (1996).
8. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. & Mello, C. C. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 706–811 (1998).
9. Lindbo, J. A., Silva-Rosales, L., Proebsting, W. M. & Dougherty, W. G. Induction of a highly specific antiviral state in transgenic plants: implications for regulation of gene expression and virus resistance. *Plant Cell* **5**, 1749–1759 (1993).
10. Jorgensen, R. A., Atkinson, R. G., Forster, R. L. S. & Lucas, W. J. An RNA-based information superhighway in plants. *Science* **279**, 1486–1487 (1998).
11. Jorgensen, R. A., Que, Q. & Stam, M. Do unintended transcripts contribute to sense cosuppression in plants? *Trends Genet.* **15**, 11–12 (1999).
12. Fire, A. RNA-triggered gene silencing. *Trends Genet.* **15**, 358–363 (1999).
13. Montgomery, M. K., Xu, S. & Fire, A. RNA as a target of double-strand RNA-mediated genetic interference in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **95**, 15502–15507 (1998).
14. Ketting, R. F., Haverkamp, T. H. A., van Luenen, H. G. A. M. & Plasterk, R. H. A. *mut-7* of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* **99**, 133–141 (1999).
15. Tabara, H. *et al.* The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**, 123–132 (1999).
16. Jensen, S., Gassama, M. & Heidmann, T. Taming of transposable elements by homology-dependent gene silencing. *Nature Genet.* **21**, 209–212 (1999).
17. Grant, S. R. Dissecting the mechanisms of posttranscriptional gene silencing: divide and conquer. *Cell* **96**, 303–306 (1999).
18. Francis, R., Barton, M. K., Kimble, J. & Schedl, T. *gld-1*, a tumour suppressor gene required for oocyte development in *Caenorhabditis elegans*. *Genetics* **139**, 579–606 (1995).
19. Ahmed, S. & Hodgkin, J. The MRT-2 checkpoint protein is required for germline immortality and telomere replication in *C. elegans*. *Nature* **403**, 159–164 (1999).
20. Collins, J., Saari, B. & Anderson, P. Activation of a transposable element in the germ line but not the soma of *Caenorhabditis elegans*. *Nature* **328**, 726–728 (1987).
21. Mori, I., Moerman, D. G. & Waterston, R. H. Analysis of a mutator activity necessary for germline transposition and excision of Tc1 transposable elements in *Caenorhabditis elegans*. *Genetics* **120**, 397–407 (1988).
22. van Blokland, R., van der Geest, N., Mol, J. N. M. & Kooter, J. M. Transgene-mediated suppression of chalcone synthase expression in *Petunia hybrida* results from an increase in RNA turnover. *Plant J.* **6**, 861–877 (1994).
23. Baulcombe, D. C. & English, J. J. Ectopic pairing of homologous DNA and post-transcriptional gene silencing in transgenic plants. *Curr. Opin. Biotechnol.* **7**, 173–180 (1996).
24. Mette, M. F., van der Winden, J., Matzke, M. A. & Matzke, A. J. Production of aberrant promoter transcripts contributes to methylation and silencing of unlinked homologous promoters in trans. *EMBO J.* **18**, 241–248 (1999).
25. Rountree, M. & Selker, E. U. DNA methylation inhibits elongation but not initiation of transcription in *Neurospora crassa*. *Genes Dev.* **15**, 2383–2395 (1997).
26. Ingelbrecht, I., van Houdt, H., Montagu, M. & Depicker, A. Posttranscriptional silencing of reporter transgenes in tobacco correlates with DNA methylation. *Proc. Natl Acad. Sci. USA* **91**, 10502–10506 (1994).
27. Jones, A. L., Thomas, C. L. & Maule, A. J. *De novo* methylation and co-suppression induced by a cytoplasmically replicating plant RNA virus. *EMBO J.* **17**, 6385–6393 (1998).
28. Selker, E. U. Gene silencing: repeats that count. *Cell* **97**, 157–160 (1999).
29. Wolffe, A. P. & Matzke, M. A. Epigenetics: regulation through repression. *Science* **286**, 481–486 (1999).
30. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970 (1991).

Acknowledgements

We thank F. Simmer, S. Ahmed and J. Hodgkin for sharing unpublished results and providing materials; T. Schedl and A. Spence for providing materials; and P. Borst, H. van Luenen and S. Fischer for critical reading of the manuscript.

Correspondence and requests for materials should be addressed to R.H.A.P. (e-mail: plasterk@niob.know.nl).

Nuclear translocation and transcription regulation by the membrane-associated guanylate kinase CASK/LIN-2

Yi-Ping Hsueh, Ting-Fang Wang*, Fu-Chia Yang & Morgan Sheng

Howard Hughes Medical Institute and Department of Neurobiology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA

Membrane-associated guanylate kinases (MAGUKs) contain multiple protein-binding domains that allow them to assemble specific multiprotein complexes in particular regions of the cell^{1,2}. CASK/LIN-2, a MAGUK required for EGF receptor localization and signalling in *Caenorhabditis elegans*, contains a calmodulin-dependent protein kinase-like domain followed by PDZ, SH3 and guanylate kinase-like domains^{3–5}. In adult rat brain, CASK is concentrated at neuronal synapses and binds to the cell-surface proteins neuexin and syndecan^{6–8} and the cytoplasmic proteins Mint/LIN-10 and Veli/LIN-7 (refs 4, 9, 10). Here we report that, through its guanylate kinase domain, CASK interacts with Tbr-1, a T-box transcription factor that is involved in forebrain development^{11,12}. CASK enters the nucleus and binds to a specific DNA sequence (the T-element) in a complex with Tbr-1. CASK acts as a coactivator of Tbr-1 to induce transcription of T-element containing genes, including *reelin*, a gene that is essential for cerebrocortical development. Our findings show that a MAGUK which is usually associated with cell junctions has a transcription regulation function.

To identify binding partners for the guanylate kinase (GK) domain of CASK, we carried out a yeast two-hybrid screen of brain complementary DNA libraries, from which Tbr-1 was isolated (Fig. 1a). Tbr-1 bound to the GK domain of CASK but not the GK domain of PSD-95. By deletion analysis, the carboxy-terminal region of Tbr-1 (residues 342–681) was found to be necessary and sufficient for association with the GK domain of CASK (Fig. 1a). A biochemical association of full-length CASK and Tbr-1 was confirmed in mammalian cells. When co-expressed in COS-7 cells, Tbr-1 and CASK were readily co-precipitated by antibodies directed against either individual protein (Fig. 1b). Chapsyn-110, a member of the PSD-95 subfamily of MAGUKs, was not co-precipitated with Tbr-1 (Fig. 1b). These biochemical data support the yeast two-hybrid results (Fig. 1a), confirming a specific interaction of Tbr-1 and CASK.

If CASK interacts physiologically with Tbr-1, it might redistribute into the nucleus in a Tbr-1-dependent fashion. After transfection in COS cells, Tbr-1 immunoreactivity was concentrated in nuclei (Fig. 2a). CASK, when expressed by itself, was cytoplasmic and relatively excluded from nuclei of COS-7 cells (Fig. 2b). Upon co-expression with Tbr-1, however, CASK redistributed into nuclei where it co-localized with Tbr-1 (Fig. 2c). This nuclear translocation is specific for CASK, because Tbr-1 did not cause the nuclear localization of other MAGUK proteins, for example, chapsyn-110 (Fig. 2c, h) and PSD-95 (Fig. 2d, i). A deletion mutant of CASK lacking the GK domain remained predominantly cytoplasmic even in the presence of nuclear Tbr-1 (Fig. 2f). Moreover, a mutant of Tbr-1 lacking the CASK-binding C-terminal region was unable to redistribute CASK into the nucleus, even though this Tbr-1 mutant itself targeted to nuclei (Fig. 2g). The nuclear localization of CASK

* Present address: Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

therefore depends on the interaction between the C-terminal region of Tbr-1 and the GK domain of CASK.

When co-transfected with haemagglutinin A (HA)-tagged Tbr-1 into cultured hippocampal neurons, CASK similarly accumulated in the nucleus with Tbr-1 (Fig. 2k). Transfection of HA-Tbr-1 alone caused endogenous CASK to partition into neuronal nuclei (Fig. 2j). Thus Tbr-1-dependent nuclear translocation of CASK occurs in neurons as well as heterologous cells.

We considered whether the Tbr-1-dependent nuclear localization of CASK could be influenced by binding partners of CASK that are usually localized to the cell surface, such as the syndecan family of transmembrane heparan sulphate proteoglycans^{7,13}. When syndecan-3 was co-expressed with Tbr-1 and CASK in COS cells, CASK no longer accumulated specifically in the nucleus but instead colocalized with syndecan-3 in a perinuclear endoplasmic reticulum (ER)-like pattern in the cytoplasm (compare Fig. 2l, middle, with 2e, left), even though Tbr-1 was concentrated in the nucleus of the same cell (Fig. 2l, right). These data suggest that the subcellular distribution of CASK may be regulated by the balance of its various binding partners in different subcellular compartments.

Does the Tbr-1-CASK complex bind to a specific DNA regulatory element? As the T-box of Tbr-1 is closely related to the T-box of Brachyury, we first examined whether Tbr-1 can recognize the 'T-element', the DNA target sequence for Brachyury¹⁴. In gel-mobility shift assays, nuclear extracts purified from COS cells expressing Tbr-1 caused a retardation of the T-element oligonucleotide probe (Fig. 3a, lane 3, band I). Peptide antibodies raised against either the carboxy or amino terminus of Tbr-1 (TBR-C and TBR-N) caused a supershift of this band, indicating that the retarded complex contains Tbr-1 (Fig. 3a, lanes 4 and 5, band II). Competition controls with unlabelled oligonucleotides confirmed the specificity of T-element binding of Tbr-1 (data not shown).

We tested whether CASK can associate with Tbr-1 in a DNA-binding complex in two ways. First, the GK domain of CASK was

added as a purified glutathione S-transferase (GST) fusion protein to Tbr-1 nuclear extracts in the presence of the T-element probe. Addition of GK domain of CASK caused a further retardation of the Tbr-1-DNA complex in a dose-dependent manner (Fig. 3b, lanes 12–14, arrow), whereas GST alone had no effect (lane 11). TBR-N antibodies supershifted this GST-CASK-Tbr-1-T-element complex (Fig. 3b, lane 15, asterisk), confirming that this complex contains Tbr-1. Second, nuclear extracts were prepared from COS cells co-transfected with CASK and Tbr-1 (Fig. 3c). These CASK-Tbr-1 nuclear extracts shifted the T-element oligo to a major band similarly to Tbr-1 alone (lane 20, band 'I'), suggesting that Tbr-1 is in excess in these nuclear extracts or that the CASK-Tbr-1-DNA complex migrates similarly to the Tbr-1-DNA complex. This complex could be supershifted with TBR-N or TBR-C antibodies, as expected (lanes 21, 22, band II). Significantly, antibodies to CASK also caused a similar, albeit weaker, supershift in Tbr-1-CASK nuclear extract (lane 23), which was not seen in nuclear extracts expressing Tbr-1 alone (lane 19), or Tbr-1 and CASKΔGK (lane 28). Thus, CASK is present in a T-element-binding complex, and formation of this complex requires the GK domain of CASK. Notably, when antibodies to both CASK and Tbr-1 are added to CASK-Tbr-1 nuclear extracts, a new 'supershifted' T-element complex was formed that barely entered the gel (lanes 24–25, band III). This slow-migrating complex, which presumably contained T-element, Tbr-1, and CASK plus antibodies to Tbr-1 and CASK, was not seen with the Tbr-1-CASKΔGK nuclear extract (Fig. 3c, lanes 26–27), indicating that the GK domain is required for CASK to associate with Tbr-1 and the T-element DNA. Collectively, the gel-retardation data show that CASK forms a GK domain-dependent complex with Tbr-1 that binds to the T-element DNA sequence.

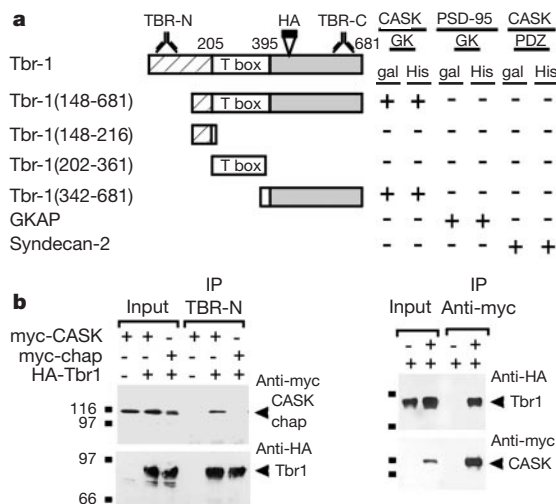


Figure 1 Specific interaction of CASK and Tbr-1. **a**, Yeast two-hybrid analysis of CASK/Tbr-1 interaction. Full-length Tbr-1 is shown at the top. Tbr-1 cDNA isolated from the yeast two-hybrid screen (residues 148–681) and other deletion constructs are aligned beneath. Interaction with the GK domain (residues 684–909) of CASK, the GK domain of PSD-95 and the PDZ domain of CASK are summarized, based on induction of reporter genes β -galactosidase (gal) and HIS3 (His)²¹. GKAP and syndecan-2 were used as positive controls for interaction with the GK domain of PSD-95 and PDZ domain of CASK, respectively^{7,22}. **b**, Co-immunoprecipitation of CASK and Tbr-1 from COS-7 cells. Cells were co-transfected with various combinations of Myc-tagged CASK, HA-tagged Tbr-1, and Myc-tagged chapsyn-110, and extracts were immunoprecipitated (IP) with TBR-N or anti-Myc antibodies, as indicated. Immunoprecipitates were immunoblotted using Myc and HA antibodies, as indicated. CASK and chapsyn-110 (chap) co-migrate on SDS-PAGE. Input lanes contain 5% of extract used for immunoprecipitation.

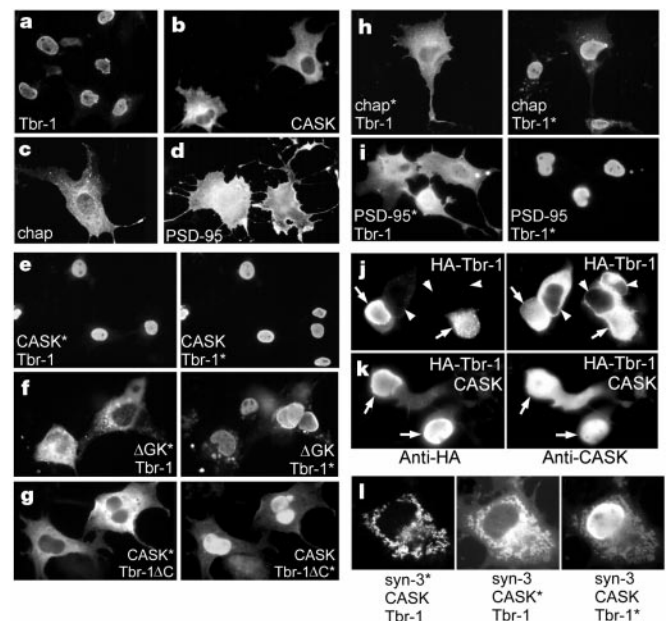


Figure 2 Tbr-1-dependent nuclear translocation of CASK. COS-7 cells were transfected with single cDNAs (**a–d**) or co-transfected with two or three different cDNAs (**e–l**), as identified in each panel. The subcellular distribution of transfected proteins was visualized by indirect immunofluorescence. For co-transfected cells, we used double-labelling (**e–l**) and triple labelling (**j**). Each group of images (in **e–l**) represents the same field viewed with FITC, Cy3 or Cy5 filters; the protein that is specifically visualized is marked by an asterisk. ΔGK, CASK mutant deleted for the GK domain (residues 708–909); Tbr-1ΔC, a deletion mutant of Tbr-1 lacking the C-terminal region (residues 401–681); chap, chapsyn-110; syn-3, syndecan-3. Similar results were obtained in HEK293 cells (not shown). **j**, **k**, double staining for HA-Tbr-1 and CASK in cultured hippocampal neurons transfected with HA-Tbr-1 alone (**j**) or both HA-Tbr-1 and CASK (**k**). Left images in **j** and **k** are stained for HA-Tbr-1; right images for CASK. Arrows indicate nuclei of transfected neurons; arrowheads point to nuclei of untransfected cells.

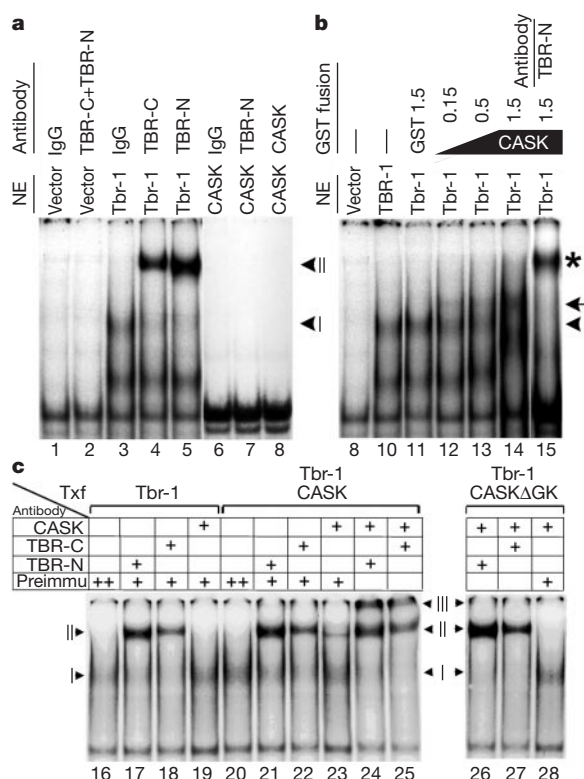


Figure 3 Tbr-1–CASK complex binds to T-element DNA. **a**, Tbr-1 binds the T-element in gel-mobility shift assays. Labelled T-element oligonucleotide was mixed with nuclear extract (NE) from COS-7 cells transfected with Tbr-1, CASK or vector control. In addition, protein-A-purified TBR-N, TBR-C and CASK antibodies, or control non-immune antibodies (IgG), were added to supershift Tbr-1-containing DNA complexes. **b**, GK domain of CASK supershifts the Tbr-1–DNA complex. Increasing amounts (0.15, 0.5 and 1.5 μ g) of purified GST-fusion protein containing the GK domain of CASK were added to Tbr-1 nuclear extract and the mixtures were analysed by T-element gel-shift assay. GST alone (1.5 μ g) was used as a negative control. **c**, Ternary complex of full-length CASK, Tbr-1 and T-element. Nuclear extracts from COS-7 cells transfected (Txf) with Tbr-1 alone or in combination with CASK or CASK Δ GK, as indicated, were used in T-element gel-shift assays. Each reaction contained anti-CASK, anti-Tbr-1 or pre-immune IgG, as indicated. The total amount of antibody was equalized by addition of pre-immune IgG. Retarded protein–DNA complexes are labelled I, II and III.

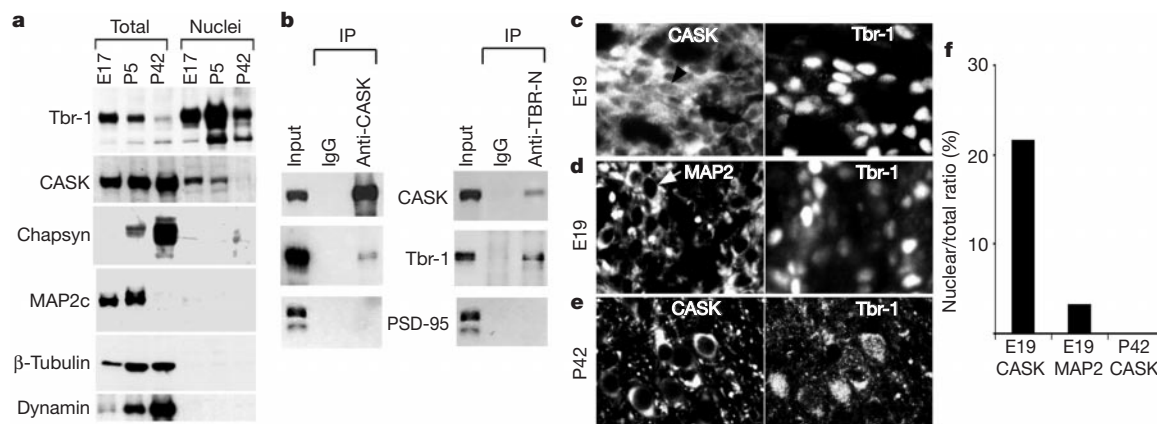


Figure 4 CASK is found with Tbr-1 in nuclei of neurons in embryonic rat cerebral cortex. **a**, Total homogenate and purified nuclei from cortex at different ages (E17, P5, P42) were immunoblotted for various proteins, as indicated. Ten micrograms of protein from each fraction, except P5 nuclei, was applied to SDS–PAGE. Because of an error in quantitation of protein, the P5 nuclei lane actually contained ~25 μ g of protein. **b**, Co-immunoprecipitation of CASK and Tbr-1 from embryonic rat cortex. Triton-X-100 extract of E20 cortex was immunoprecipitated with TBR-N or CASK antibodies, or non-immune IgG, as indicated. The precipitates were immunoblotted for CASK, Tbr-1 and

Tbr-1 is more highly expressed in embryonic brain¹¹ (Fig. 4a), therefore the interaction between Tbr-1 and CASK may be particularly important in developing brain. Expression of CASK is roughly constant from embryonic day E17 to postnatal day P42. A significant fraction of CASK is present in nuclei purified from E17 and P5 cerebral cortex (Fig. 4a). Roughly 10–20% of CASK was present in nuclei at E17, and this percentage declined with development; at P42, nuclear CASK was barely detectable by immunoblotting. Several cytoplasmic proteins (chapsyn-110, β -tubulin, dynamin, MAP2c) were not detected in the purified nuclear fractions; thus, the CASK signal in E17 and P5 nuclei is unlikely to represent cytoplasmic contamination.

To examine whether Tbr-1 and CASK coexist in the same nuclei of prenatal neocortex, we carried out double-label immunofluorescence confocal microscopy on E19 brain sections. As expected, Tbr-1 was localized in the nuclei of neurons (Fig. 4c, right). CASK was diffusely distributed, mainly in a cytoplasmic pattern. However, there was weak but significant immunoreactivity for CASK in nuclei (Fig. 4c, left), resulting in a relatively ‘filled-in’ appearance of the nuclei compared with MAP2 staining at E19 (Fig. 4d, left) or CASK staining at P42 (Fig. 4e, left), both of which were excluded from nuclei. We used quantitative immunofluorescence to estimate the fraction of CASK in nuclei (Fig. 4f). About 20% of CASK immunofluorescence was localized to nuclei in E19 rat brain, consistent with immunoblotting data (Fig. 4a). In contrast, nuclear immunoreactivity of MAP2 in E19 rat brain represented only ~3% of the total, and nuclear CASK in P42 brain was undetectable (Fig. 4f).

We could co-immunoprecipitate Tbr-1 (but not PSD-95) from embryonic cortex using CASK antibodies (Fig. 4b, left), and vice versa (Fig. 4b, right), providing more evidence for the association of CASK and Tbr-1 *in vivo*. We were unable to co-immunoprecipitate CASK and Tbr-1 from P42 brain when expression of Tbr-1 was low.

There is a change in the subcellular distribution of Tbr-1 and CASK during cortical development (Fig. 4c, e). In P42 cortex, CASK is found in a punctate pattern in the neuropil, with additional cytoplasmic staining in neuronal cell bodies (Fig. 4e, left). At this stage, Tbr-1 is still present in the nucleus of neurons, but no longer restricted to nuclei as it was at E19. Instead, Tbr-1 distribution extends into the cytoplasm of P42 neurons (Fig. 4e, right), overlapping considerably with CASK outside of the nucleus.

What is the functional consequence of CASK binding on the putative transcriptional activity of Tbr-1? First, we tested whether

Tbr-1 affects the expression of a luciferase reporter gene placed under the control of an upstream T-element (T-luc) in heterologous cells. Consistent with earlier studies¹⁴, Brachyury strongly stimulated T-luc expression (~20-fold, Fig. 5a). In comparison, Tbr-1 induction of T-luc expression was weak (~two–threefold relative to vector control). Co-expression of CASK, however, enhanced the transcription stimulation by Tbr-1 to ~10-fold (Fig. 5a). A CASK mutant with deleted GK domain (CASKΔGK) was unable to stimulate Tbr-1 activation of T-luc, and wild-type CASK had no effect on the transcription activity of Brachyury. In addition, CASK itself had no significant effect on T-luciferase expression (Fig. 5a). We conclude that CASK specifically stimulates transcriptional activity of Tbr-1 by GK-dependent binding to Tbr-1.

A fusion protein of the GAL-4 DNA-binding domain and the C-terminal region of Tbr-1 (GAL-4–Tbr-1C) had no effect on the GAL-4-luciferase reporter gene (Fig. 5b), suggesting that the CASK-binding domain of Tbr-1 itself has no transcriptional activity. However, co-expression of CASK with GAL-4–Tbr-1C greatly stimulated GAL4-luc expression (Fig. 5b), confirming that CASK can act as a ‘coactivator’ of Tbr-1. CASK mutants individually lacking calmodulin-dependent protein kinase (CaMK), Veli/LIN-7-binding, PDZ or SH3 domains were still able to coactivate GAL4-luc transcription in the presence of GAL-4–Tbr-1C, but deleting the GK domain abolished this activity (Fig. 5b). Moreover, the GK domain alone was as effective as full-length CASK (Fig. 5b); thus, the GK domain of CASK is necessary and sufficient for the coactivator activity of CASK on Tbr-1.

We also tested the transcriptional effects of Tbr-1 and CASK in neurons. Overexpression of Tbr-1 in cultured neurons stimulated expression of the co-transfected T-luc reporter gene (Fig. 6a). Notably, in contrast to the data from COS cells, transfection of CASK alone was able to induce T-luc expression in cultured neurons (>10-fold; Fig. 6a). This result can be explained by the fact that Tbr-1 is endogenously expressed in these neurons (data not shown). Co-transfection of both CASK and Tbr-1 further enhanced T-luc

activity (Fig. 6a). The stimulatory effect of CASK on T-element-directed transcription therefore occurs in neurons as well as heterologous cells.

Tbr-1 expression is highly restricted to cerebral cortex, particularly at prenatal stages¹¹, and is undetectable in lung, heart, liver and kidney of E20 rat embryos (data not shown), suggesting that Tbr-1 may have a role in cerebral cortex development. Consistent with this idea, Tbr-1 knockout mice show a developmental phenotype similar to Reeler mice, which is characterized by abnormal lamination of cerebral cortex¹². We noticed that the 5' upstream region of the reelin gene (which is mutated in Reeler mice) contains two half-palindromes of the T-element, and we proposed that reelin gene transcription may be controlled by Tbr-1–CASK. To test this idea, we determined in transfected neurons the activity of Tbr-1 and CASK on a luciferase reporter gene under the control of the reelin 5' upstream region (Reelin-luc). Tbr-1 itself stimulated Reelin-luc expression only modestly (~twofold relative to control vector; Fig. 6b), perhaps because Tbr-1 binds weakly to half-palindromes of the T-element. Overexpression of CASK by itself greatly induced expression of Reelin-luc in neurons (~16 fold; Fig. 6b). Moreover, when co-transfected with both CASK and Tbr-1, expression of Reelin-luc was further enhanced (~28 fold; Fig. 6b). In addition to confirming the transcriptional stimulatory activity of CASK, these results indicate that the Tbr-1–CASK complex can positively regulate a natural gene promoter in neurons and suggest that reelin is an endogenous target gene of Tbr-1–CASK.

We have reported the first example, to our knowledge, of a MAGUK translocating into the nucleus, interacting with a defined transcription factor and regulating transcription. Like other MAGUKs, CASK has been mainly characterized as a membrane-associated scaffold protein, involved in the assembly of specific protein complexes at sites of cell contact. By also regulating transcription, CASK invites analogies with β-catenin, a cadherin-binding protein that is associated with adherens junctions which can also function as a nuclear mediator by binding to Tcf/LEF

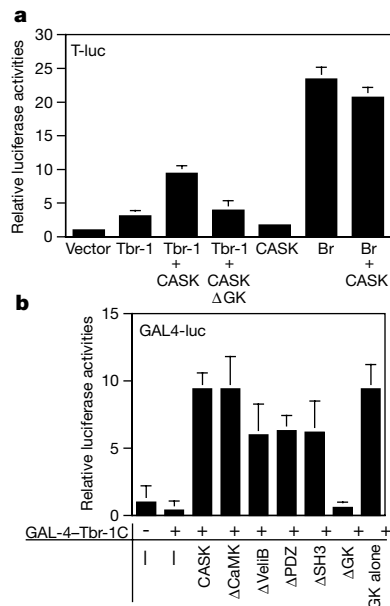


Figure 5 CASK stimulates transcriptional activity of Tbr-1 in COS-7 cells. **a**, Luciferase reporter gene under the control of T-element (T-luc) was co-transfected with Tbr-1, CASK, CASKΔGK, Brachyury (Br) or vector control, as indicated. **b**, GK domain of CASK is necessary and sufficient for coactivation effect of CASK. GAL-4-luciferase reporter gene (GAL-4-luc) was co-transfected with control vector, GAL-4-Tbr-1C, plus various CASK constructs, as indicated. Δ indicates the specific single-domain deletions of CASK. CASK, full-length CASK; GK alone, CASK GK domain alone. Luciferase activities are normalized to vector control. Error bars show standard deviations ($n = 6$).

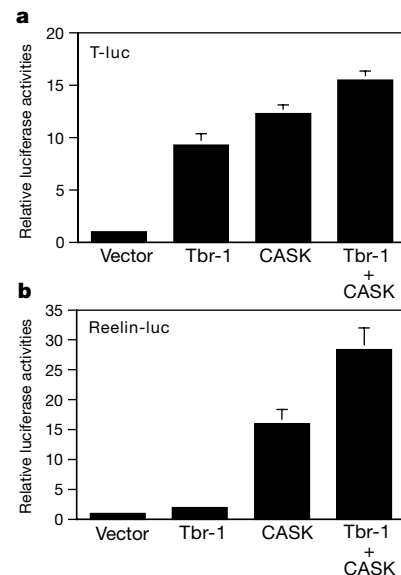


Figure 6 Tbr-1/CASK complex activates gene expression in cultured neurons. **a**, CASK–Tbr-1 stimulates T-luc expression. Hippocampal neurons (3DIV) were co-transfected with T-luc reporter and Tbr-1 and/or CASK, as indicated. **b**, CASK–Tbr-1 complex activates expression of luciferase reporter gene under the control of Reelin gene upstream region (Reelin-luc). Reelin-luc was co-transfected with Tbr-1 and/or CASK into cultured hippocampal neurons, as indicated. Error bars show standard deviations ($n = 3$).

transcription factors^{15,16}. Another MAGUK (ZO-1) has been shown to shift from the tight junction to the nucleus in a cell contact-regulated manner, but the mechanism of this redistribution is unknown¹⁷.

Although we detected nuclear CASK only in embryonic neurons, it remains possible that CASK translocates from cell junction to nucleus in mature neurons. Such a translocation may be highly regulated and/or of small magnitude, thus preventing the detection of CASK in the nuclei of mature neurons. In a heterologous system, we have found that a transmembrane ligand of CASK (syndecan-3) can shift the balance of CASK distribution between nucleus and cytoplasm (Fig. 2). Studying the dynamic regulation of CASK interactions with neuroligins and syndecans may therefore be key to understanding the potential involvement of CASK in cell surface-to-nucleus signalling. □

Methods

Plasmid constructs

Tbr-1 coding sequence was RT-PCR amplified from mouse brain cDNA and subcloned into eukaryotic expression vector GW1-CMV (British Biotechnology). The HA epitope was inserted into a unique *Ascl* site (between residues 479–480) of Tbr-1. To construct T-luc, two repeats of T element (AATTCACACCTAGGTGTGAAATT) were subcloned into reporter plasmid ΔERSV90-Luc, which contains the basal promoter of RSV fused to luciferase. Reelin-luc contains ~1 kb of 5' upstream region of the reelin gene (GenBank accession number AC002067, nucleotides 3,700–4,620). GAL-4-Tbr-1C contains residues 342–681 of Tbr-1.

Transfection, immunoprecipitation, immunohistochemistry, luciferase assay

Transfection (using Lipofectamine), immunocytochemistry and immunoprecipitation of COS-7 cells were done as described¹⁸. For immunohistochemistry, rat brain sections were processed by indirect immunofluorescence as described⁷, with the following modifications. E19 rat brains were fixed with 4% formaldehyde in PBS at 4 °C overnight, and cryoprotected with 20% sucrose in PBS. Frozen sections (25-μm) were collected on slides, air dried for at least 30 min, and permeabilized with methanol at –20 °C for 20 min. Results were viewed with an Axioskop microscope or a Bio-Rad MRC-1000 confocal microscope. For luciferase assays, COS-7 cells and cultured hippocampal neurons were harvested 18–20 h after transfection. Luciferase activity was measured in equal amounts of cell lysate protein using the Luciferase assay system (Promega).

Subcellular fractionation from rat brain

Purification of nuclei from rat cerebral cortex was done as described¹⁹.

DNA-binding gel-mobility shift assay

To prepare nuclear extracts, transfected COS-7 cells were harvested, quickly washed with hypotonic buffer (20 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM MgCl₂, 0.5 mM DTT), and resuspended in a volume of hypotonic buffer about five times the packed cell volume. Cells were swollen at 4 °C for 10 min and broken by dounce homogenization with a type B pestle. Nuclei were collected and resuspended in buffer B (20 mM HEPES pH 7.9, 1.5 mM MgCl₂, 0.5 mM DTT, 20% glycerol) containing 0.3 M KCl. The nuclei were further extracted with gentle mixing at 4 °C for 30 min, after which nuclear debris was removed by centrifugation in a microcentrifuge at 14,000g for 10 min. Nuclear extracts were diluted by addition of two volumes of buffer B (reducing the concentration of KCl to 0.1 M) before use. Gel mobility shift assays were done as described²⁰. The T-element oligonucleotides were annealed and labelled by T4 polynucleotide kinase with [γ-³²P]ATP.

Received 10 November 1999; accepted 19 January 2000.

- Craven, S. E. & Bredt, D. S. PDZ proteins organize synaptic signaling pathways. *Cell* **93**, 495–498 (1998).
- Anderson, J. M. Cell signalling: MAGUK magic. *Curr. Biol.* **6**, 382–384 (1996).
- Kim, S. K. Polarized signaling: basolateral receptor localization in epithelial cells by PDZ-containing proteins. *Curr. Opin. Cell Biol.* **9**, 853–859 (1997).
- Kaech, S. M., Whitfield, C. W. & Kim, S. K. The LIN-2/LIN-7/LIN-10 complex mediates basolateral membrane localization of the *C. elegans* EGF receptor LET-23 in vulval epithelial cells. *Cell* **94**, 761–771 (1998).
- Bredt, D. S. Sorting out genes that regulate epithelial and neuronal polarity. *Cell* **94**, 691–694 (1998).
- Hata, Y., Butz, S. & Südhof, T. C. CASK: a novel *dlg*/PSD95 homolog with an N-terminal calmodulin-dependent protein kinase domain identified by interaction with neuroligins. *J. Neurosci.* **16**, 2488–2494 (1996).
- Hsueh, Y.-P. *et al.* Direct interaction of CASK/LIN-2 and syndecan heparan sulfate proteoglycan and their overlapping distribution in neuronal synapses. *J. Cell Biol.* **142**, 139–151 (1998).
- Cohen, A. R. *et al.* Human Cask/Lin-2 binds syndecan-2 and protein 4.1 and localizes to the basolateral membrane of epithelial cells. *J. Cell Biol.* **142**, 129–138 (1998).
- Borg, J. P. *et al.* Identification of an evolutionarily conserved heterotrimeric protein complex involved in protein targeting. *J. Biol. Chem.* **273**, 31633–31636 (1998).
- Butz, S., Okamoto, M. & Südhof, T. C. A tripartite protein complex with the potential to couple

- synaptic vesicle exocytosis to cell adhesion in brain. *Cell* **94**, 773–782 (1998).
- Bulfone, A. *et al.* T-brain-1: a homolog of *Brachyury* whose expression defines molecularly distinct domains within the cerebral cortex. *Neuron* **15**, 63–78 (1995).
- Hevner, R. F., Shi, L., LeWinter, R. & Rubenstein, J. L. R. Cortical phenotype of Tbr-1 mutant mice: laminar inversion and decreased expression of Reelin. *Soc. Neurosci. Abstr.* **25**, 502 (1999).
- Hsueh, Y. P. & Sheng, M. Regulated expression and subcellular localization of syndecan heparan sulfate proteoglycans and the syndecan-binding protein CASK/LIN-2 during rat brain development. *J. Neurosci.* **19**, 7415–7425 (1999).
- Kispert, A., Koschorz, B. & Herrmann, B. G. The T protein encoded by *Brachyury* is a tissue-specific transcription factor. *EMBO J.* **14**, 4763–4772 (1995).
- Miller, J. R. & Moon, R. T. Signal transduction through beta-catenin and specification of cell fate during embryogenesis. *Genes Dev.* **10**, 2527–2539 (1996).
- Peifer, M. β-catenin as oncogene: the smoking gun. *Science* **275**, 1752–1753 (1997).
- Gottardi, C. J., Arpin, M., Fanning, A. S. & Louvard, D. The junction-associated protein, zonula occludens-1, localizes to the nucleus before the maturation and during the remodeling of cell-cell contacts. *Proc. Natl Acad. Sci. USA* **93**, 10779–10784 (1996).
- Hsueh, Y. -P. & Sheng, M. Requirement of N-terminal cysteines of PSD-95 for PSD-95 multimerization and ternary complex formation, but not for binding to potassium channel Kv1.4. *J. Biol. Chem.* **274**, 532–536 (1999).
- Gorski, K., Carneiro, M. & Schibler, U. Tissue-specific in vitro transcription from the mouse albumin promoter. *Cell* **47**, 767–776 (1986).
- Hsueh, Y.-P. & Lai, M.-Z. Overexpression of activation transcriptional factor 1 in lymphomas and in activated lymphocytes. *J. Immunol.* **154**, 5675–5683 (1995).
- Niethammer, M. & Sheng, M. Identification of ion channel-associated proteins using the yeast two-hybrid system. *Methods Enzymol.* **293**, 104–122 (1998).
- Kim, E. *et al.* GKAP, a novel synaptic protein that interacts with the guanylate kinase-like domain of the PSD-95/SAP90 family of channel clustering molecules. *J. Cell Biol.* **136**, 669–678 (1997).

Acknowledgements

We thank J. Trimmer and L. Buchwalder for CASK monoclonal antibodies; C. Konradi for luciferase reporter plasmid ΔERSV90-luc; A. Kispert for *Brachyury* (T) cDNA; M. Greenberg and J. Kornhauser for GAL-4 reporter vectors; L.-H. Tsai and Y. Kwon for advice on embryonic brain immunohistochemistry. M.S. is Assistant Investigator of the Howard Hughes Medical Institute. This work was supported by the NIH (M.S.) and a National Neurofibromatosis Foundation Young Investigator Award (Y.-P.H.).

Correspondence and requests for materials should be addressed to M.S. (e-mail: sheng@helix.mgh.harvard.edu).

Phosphorylation of CPE binding factor by Eg2 regulates translation of *c-mos* mRNA

Raul Mendez*, Laura E. Hake*†, Thorkell Andresson‡, Laurie E. Littlepage‡, Joan V. Ruderman‡ & Joel D. Richter*

* Department of Molecular Genetics and Microbiology, University of Massachusetts Medical Center, Worcester, Massachusetts 01655, USA

‡ Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

Full-grown *Xenopus* oocytes arrest at the G2/M border of meiosis I. Progesterone breaks this arrest, leading to the resumption of the meiotic cell cycles and maturation of the oocyte into a fertilizable egg. In these oocytes, progesterone interacts with an unidentified surface-associated receptor, which induces a non-transcriptional signalling pathway that stimulates the translation of dormant *c-mos* messenger RNA. Mos, a mitogen-activated protein (MAP) kinase kinase kinase, indirectly activates MAP kinase, which in turn leads to oocyte maturation. The translational recruitment of *c-mos* and several other mRNAs is regulated by cytoplasmic polyadenylation, a process that requires two 3' untranslated regions, the cytoplasmic polyadenylation element (CPE) and the polyadenylation hexanucleotide AAUAAA^{1–4}. Although the signalling events that trigger *c-mos* mRNA polyadenylation and translation are unclear, they probably involve the activation of

† Present address: Department of Biology, Boston College, Chestnut Hill, Massachusetts 02167, USA.

CPEB, the CPE binding factor^{5,6}. Here we show that an early site-specific phosphorylation of CPEB is essential for the polyadenylation of *c-mos* mRNA and its subsequent translation, and for oocyte maturation. In addition, we show that this selective, early phosphorylation of CPEB is catalysed by Eg2, a member of the Aurora family of serine/threonine protein kinases.

Previous work showed that CPEB is phosphorylated at a time that is commensurate with the polyadenylation of several mRNAs such as those encoding histone B4 and cyclin B1 (refs 5, 7). However,

adenylation of those mRNAs has subsequently been shown to occur late in oocyte maturation and to depend on the earlier adenylation and translation of *c-mos* mRNA^{8,9}. In addition, *cdc2*-mediated phosphorylation of CPEB^{5,7}, which is detected by a mobility shift in the protein, is dependent on Mos synthesis^{5,7,8,10}. Therefore the late-stage phosphorylation of CPEB cannot be involved in the activation of *c-mos* mRNA translation.

Because many phosphorylations of substrate proteins do not induce mobility shifts in SDS-polyacrylamide gel electrophoresis

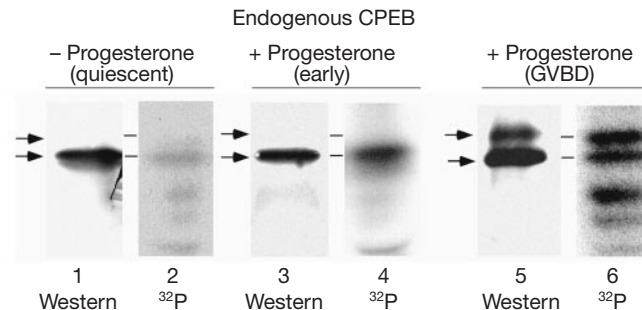


Figure 1 CPEB is phosphorylated early after progesterone stimulation and hyperphosphorylated at maturation. Stage VI *Xenopus* oocytes were metabolically labelled with ³²P (ref. 5), stimulated with progesterone and collected at zero (- progesterone, quiescent), 2 (+ progesterone, early) or 6 (+ progesterone, GVBD) hours. After immunoprecipitation⁵,

CPEB was resolved by SDS-PAGE and detected by western blotting (lanes 1, 3, 5); the phosphorylation state was analysed by autoradiography (lanes 2, 4, 6). CPEB positions (shifted and unshifted) are indicated by the arrows.

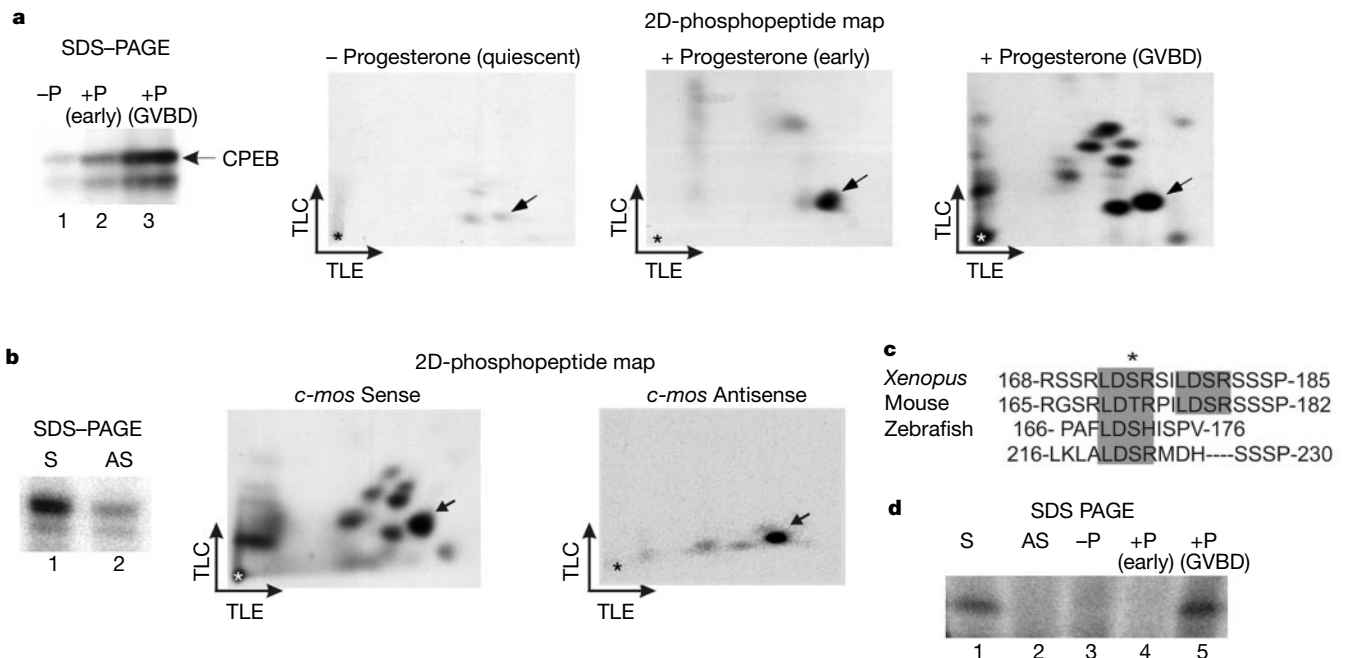


Figure 2 Phosphorylation of Pp1 occurs early after progesterone stimulation and does not require Mos synthesis. **a**, Recombinant His-tagged CPEB was phosphorylated *in vitro* with extracts from immature oocytes (- P) or from oocytes stimulated with progesterone for 2 (early) or 6 (GVBD) hours, when maturation occurred. CPEB was then purified, resolved by SDS-PAGE, western blotted and digested with trypsin; the resultant phospho-peptides were analysed by 2D peptide mapping²⁴ and autoradiography. Arrows indicate the directions of the chromatographic (TLC; phospho-chromatography buffer, n-butanol 37.5%, pyridine 25%, acetic acid 7.5%) and electrophoretic (TLE; pH 1.9) migrations. The star indicates the origin. **b**, Recombinant His-tagged CPEB was phosphorylated with extracts from stage VI oocytes that were injected with *c-mos* sense (S) or antisense (AS) oligodeoxynucleotides⁸ and stimulated with progesterone for 6 hours. CPEB was treated as in **a** to obtain phospho-peptides. The phospho-peptide Pp1 (arrow) was extracted,

purified by HPLC and sequenced by Edman degradation, revealing a tryptic peptide (LDSR) corresponding to residues 172-175 of CPEB, where Ser 174 is the phosphorylated residue. **c**, Sequence comparison of CPEB from *Xenopus*, mouse and zebrafish. The LDSR/TR or LDSH motif in the different proteins is boxed, and the phosphoserine is identified by a star. **d**, A recombinant His-tagged CPEB variant in which Ser 174 and Ser 180 were mutated to alanine (CPEB-AA) was phosphorylated *in vitro* with extracts from stage VI oocytes that had been injected with sense (S) or antisense (AS) oligodeoxynucleotides for *c-mos* mRNA⁸ and stimulated with progesterone for 6 hours. The protein was also phosphorylated with extracts from immature oocytes (-P) or from oocytes stimulated with progesterone for 2 (early) or 6 (GVBD) hours when maturation occurred. CPEB was purified and analysed by SDS-PAGE and autoradiography.

(PAGE), it seemed possible that an early CPEB phosphorylation that precedes *c-mos* mRNA adenylation and translation and *cdc2* activation might have been missed in previous experiments. To look for an early CPEB phosphorylation, we incubated quiescent or progesterone-stimulated oocytes with ^{32}P and carried out CPEB immunoprecipitation, western blotting and autoradiography (Fig. 1). As reported previously^{5,7,8}, western blots show that CPEB was present in immature oocytes as a single species (lane 1), whereas at maturation (that is, germinal vesicle breakdown (GVBD)) an additional slow-migrating form of the protein, which was the result of phosphorylation, was observed (lane 5). When CPEB phosphorylation was detected by *in vivo* labelling, however, CPEB showed a low level of phosphorylation even in immature oocytes (lane 2), and was heavily phosphorylated early on during maturation at a time coincident with *c-mos* mRNA polyadenylation (lane 4). This early phosphorylation did not result in an altered mobility of CPEB, in contrast to when phosphorylation occurs late in maturation (lane 6), showing that CPEB undergoes multiple, time-dependent phosphorylations during maturation.

To examine the sites of CPEB phosphorylation at early and late times after progesterone stimulation, recombinant CPEB that was

phosphorylated with extracts from unstimulated and progesterone-stimulated oocytes collected before (early) and after (GVBD) *cdc2* activation was examined by SDS-PAGE and two-dimensional phospho-peptide mapping (Fig. 2a). A single tryptic phosphopeptide (Pp1) was detected when early extracts were used; in contrast, eight phospho-peptides were detected when the GVBD extracts were used.

If early phosphorylation of CPEB is required for the translational recruitment of *c-mos* mRNA, it should precede Mos synthesis and all Mos-dependent downstream events. To examine this, oocytes were first injected with an antisense oligonucleotide to destroy *c-mos* mRNA and then stimulated with progesterone. This treatment effectively blocks Mos protein accumulation and thus the activation of both MAP kinase and *cdc2*, the CPEB mobility shift, the late polyadenylation events such as those that take place on cyclin B1 and histone B4 mRNAs, and oocyte maturation¹¹. Extracts prepared from these oocytes were still able to phosphorylate Pp1, but not the latter sites (Fig. 2b, right). These data show that progesterone induces the phosphorylation of a single CPEB peptide before Mos synthesis, and suggests that phosphorylation at this early site could be responsible for the activation of CPEB.

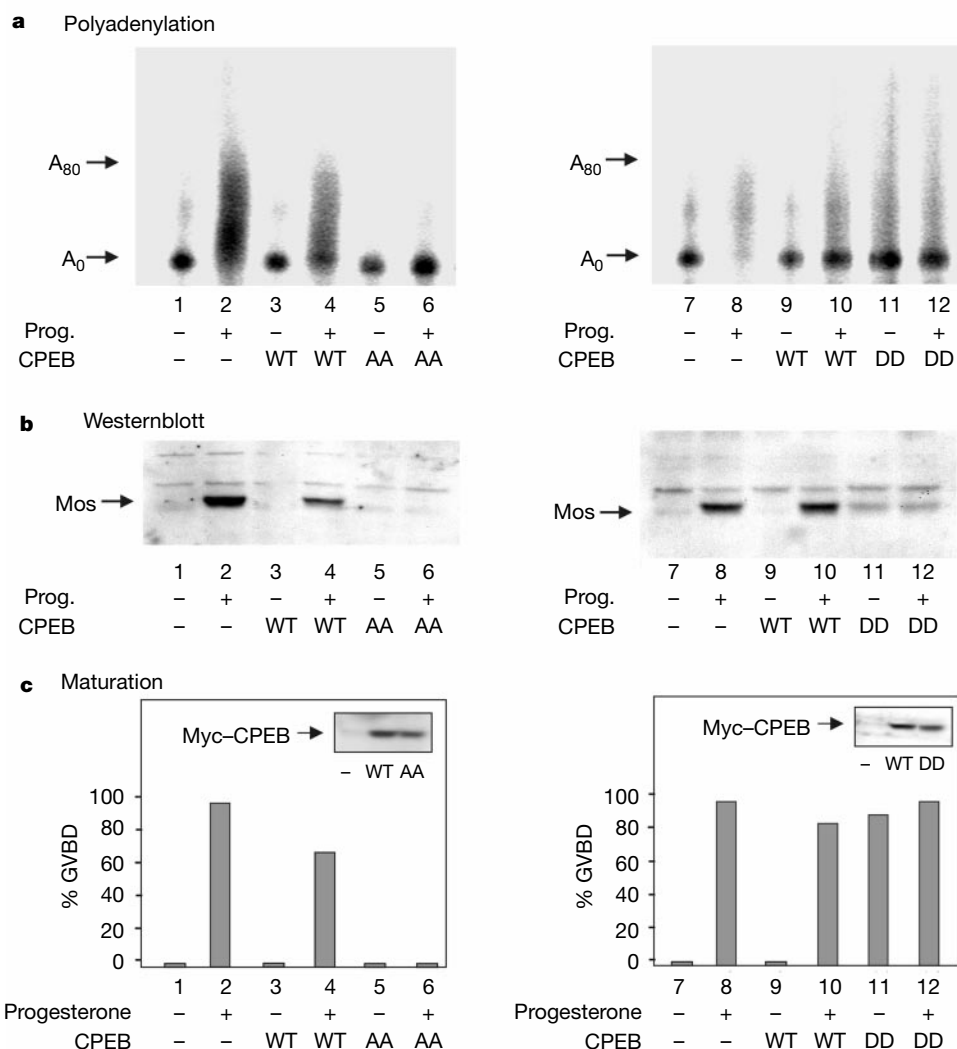


Figure 3 Phosphorylation of CPEB is required for polyadenylation-induced translation. Oocytes were injected with *in vitro* transcribed mRNAs encoding Myc-tagged wild-type CPEB (CPEB-WT), or CPEB variants in which Ser 174 and Ser 180 were mutated to either alanine (CPEB-AA) or aspartic acid (CPEB-DD). After 12 hours, all oocytes were injected a second time with a radioactive *c-mos* 3' UTR, and incubated in the absence or presence of progesterone. The oocytes were analysed for polyadenylation of the *c-mos* 3' UTR⁸ (a),

Mos synthesis by Western blotting⁸ (b), and maturation (GVBD) as detected by a white spot at the animal pole (c). The levels of expression of the injected CPEB mRNAs were analysed by western blots probed with an anti-Myc antibody (insets). The CPEB-DD mRNA injection results were somewhat variable (3 out of 6 experiments resulted in oocyte maturation), probably because of inherent variabilities in different batches of oocytes.

To identify the site(s) of Pp1 phosphorylation, this peptide was extracted from the thin-layer chromatography (TLC) plate, purified by high-performance liquid chromatography (HPLC) and sequenced. The Pp1 sequence was LDSR, which corresponds to residues 172–175 of CPEB, where Ser 174 was the phosphorylated

residue. An LDSR motif, or the related LDTR and LDSH motifs which could be functionally similar, is found in all known vertebrate CPEBs and tandemly repeated in the *Xenopus* and mouse proteins (Fig. 2c). A variant of CPEB that contained alanine substitutions for Ser 174 and Ser 180 was not phosphorylated in extracts prepared

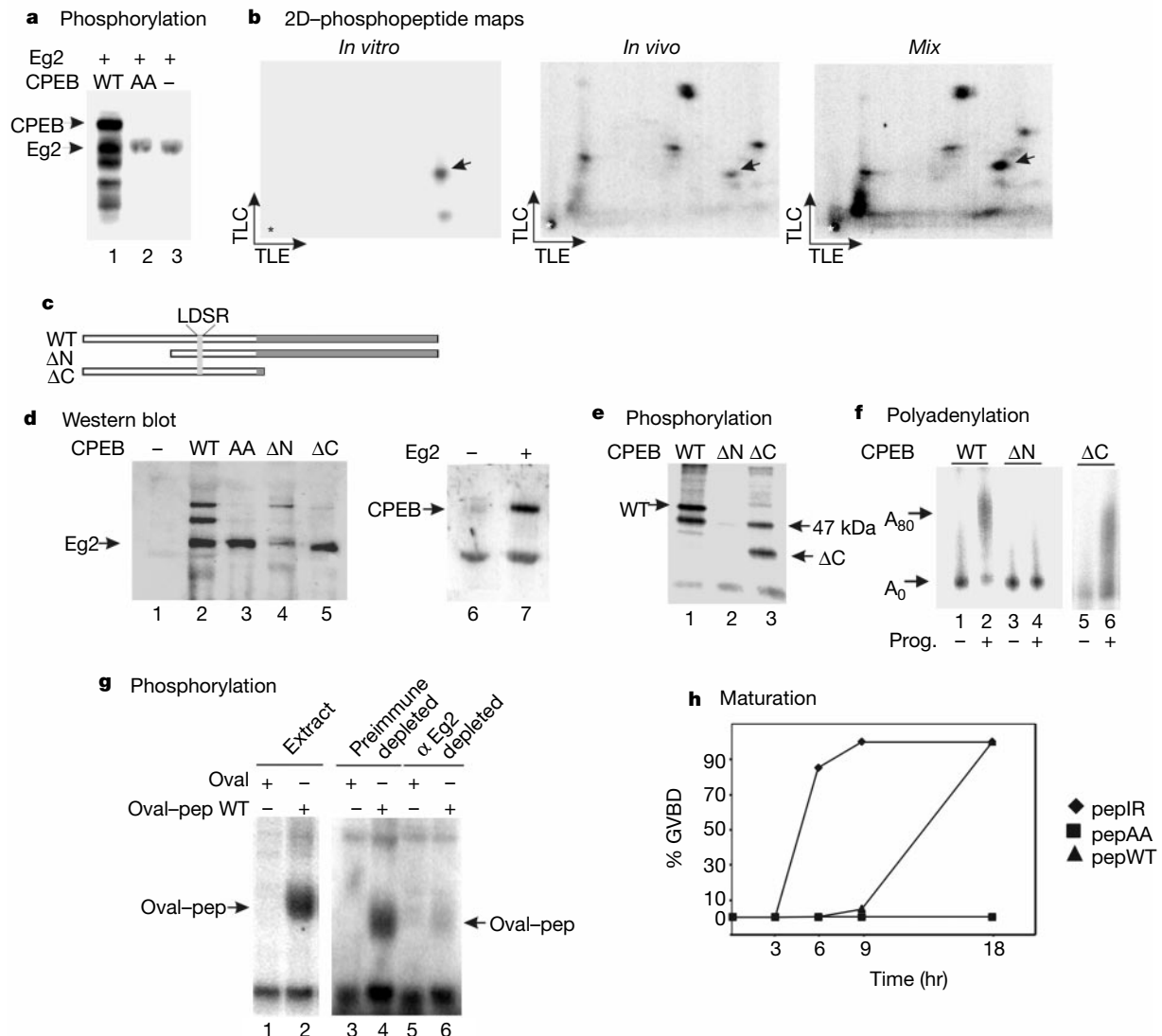


Figure 4 Eg2 phosphorylates CPEB at Ser 174. **a**, Recombinant baculovirus expressed Eg2 (ref. 12) was incubated with [γ -³²P]ATP and recombinant wild-type His-tagged CPEB (CPEB-WT), the S174A/S180A CPEB variant (CPEB-AA), or with no substrate addition (-). Reaction products were analysed by SDS-PAGE and autoradiography. **b**, Recombinant CPEB was phosphorylated with Eg2 *in vitro* in the presence of [γ -³²P]ATP (*in vitro*). Phospho-peptide maps of this CPEB, and of endogenous CPEB (*in vivo*) radiolabelled in mature (that is, GVBD) oocytes, were obtained as in Fig. 1. These phospho-peptides were also mixed before the 2D mapping (mix). Arrow denotes Pp1. **c**, Representation of wild-type (WT), N-terminally deleted (lacking residues 1–146, ΔN), and C-terminally deleted (lacking residues 290–568, ΔC) CPEB proteins. The RNA-binding domain of CPEB is shaded, and the position of the LDSR motif is indicated. **d**, These His-tagged CPEB proteins, including one with S174A and S180A mutations (AA), were incubated with mature oocyte extract for 30 min at 30 °C in phosphorylation buffer. CPEB-containing complexes were then captured on nickel–agarose, eluted with 1 M NaCl and analysed by immunoblotting with Eg2 antibody. A control with no CPEB addition was carried out in parallel (lane 1). His-tagged Eg2 (lane 7) was also incubated with mature oocyte extract, and Eg2-containing complexes were captured on nickel–agarose and probed for CPEB on a western blot. A control with no Eg2 addition was carried out in parallel (lane 6). **e**, Recombinant His-tagged CPEB variants (WT, ΔN, ΔC) were phosphorylated *in vitro* with extract from progesterone-stimulated oocytes as the kinase source, purified by nickel

chromatography, and analysed by SDS-PAGE and autoradiography. A 47K phospho-protein that co-purifies with WT and ΔC CPEB proteins, which is probably Eg2, is indicated. **f**, Oocytes injected with *in vitro* transcribed mRNAs encoding WT, ΔN and ΔC CPEB proteins and a radioactive *c-mos* 3' UTR were incubated in the absence or presence of progesterone. When the CPEB-WT mRNA injected oocytes had matured, RNA was extracted from all oocytes and the polyadenylation of the *c-mos* 3' UTR was assessed⁹. **g**, A peptide corresponding to residues 163–181 of *Xenopus* CPEB was coupled to ovalbumin (Oval-pep). This coupled peptide (lanes 2, 4, 6) or ovalbumin alone (Oval; lanes 1, 3, 5) was incubated with [γ -³²P]ATP and mature oocyte extract (lanes 1, 2), extract mock depleted with preimmune serum⁶ (lanes 3, 4) or extract immunodepleted⁶ with Eg2 antibody (lanes 5, 6). The reaction products were analysed by SDS-PAGE and autoradiography. The position of oval-pep and ovalbumin are indicated by the arrows. **h**, Oocytes injected with ovalbumin coupled to an irrelevant peptide (IR, sequence EFKLAAEADAAAAADMDKFC), a wild-type CPEB derived peptide containing the Eg2 phosphorylation site (WT, sequence RGSRLDTRPILDSRSSC) or to an alanine-substituted peptide (AA, sequence RGSRLDARPIILDARSSC) were cultured in the presence of progesterone. Oocyte maturation was then scored. Each oocyte was injected with 50 nl of a 2 mg ml⁻¹ solution of the peptides coupled to ovalbumin that contained an average of 4 mol of peptide per mol of ovalbumin.

from early or *c-mos* antisense oligodeoxynucleotide-injected oocytes, which confirms that the single phospho-peptide detected with those extracts was indeed Pp1 (Fig. 2d). Although we did not detect Ser 180 phosphorylation, which also resides in the sequence LDSR, we changed both residues to prevent a possible alternative phosphorylation event that might mask the importance of Ser 174 phosphorylation.

To assess the function of the early CPEB phosphorylation, oocytes were injected with mRNA encoding either wild-type or mutant CPEB-AA (S174A, S180A) and subsequently with a *c-mos* 3' untranslated region (UTR) reporter construct to examine polyadenylation. Messenger RNA encoding wild-type CPEB (CPEB-WT) had no effect on progesterone-induced *c-mos* RNA polyadenylation, but the mRNA encoding mutant CPEB completely prevented polyadenylation of the reporter construct (Fig. 3a, left). Moreover, injection of mRNA encoding mutant CPEB blocked both endogenous Mos synthesis (Fig. 3b, left) and oocyte maturation (Fig. 3c, left). The observation that the mutated CPEB containing the non-phosphorylatable alanine residues acts as a dominant-negative component in this assay indicates that the function of CPEB may require the phosphorylation of Ser 174 before Mos synthesis.

Because the alanine substitutions cause CPEB to act in a dominant-negative fashion, we wanted to determine whether negatively charged residues such as aspartic acid could mimic the effect of phosphorylation and create a dominant gain-of-function CPEB mutant. Accordingly, mRNA encoding wild-type or mutant (S174D, S180D; DD) CPEB was injected into oocytes, followed by the *c-mos* reporter RNA noted above. Messenger RNA encoding wild-type CPEB did not induce the polyadenylation of *c-mos* RNA (Fig. 3a, right). In contrast, mRNA encoding DD-CPEB stimulated polyadenylation even in the absence of progesterone (Fig. 3a, right). This polyadenylation was not as robust as that observed in uninjected oocytes incubated with progesterone, but it was sufficient to induce elevated levels of endogenous Mos accumulation (Fig. 3b, right) and a very high incidence of oocyte maturation (Fig. 3c, right). These effects were not further increased by the addition of progesterone. These results indicate that the early phosphorylation of CPEB at Ser 174 may be a key regulatory step and that this modification may be sufficient to trigger cytoplasmic polyadenylation and the subsequent signalling cascades that result in meiotic progression.

The serine/threonine kinase Eg2 has been identified as an early-acting component of the progesterone signalling pathway that functions upstream of *c-mos* mRNA translational activation¹². Eg2, which is related to Aurora¹³ from *Drosophila* and IPL1 (ref. 14) from *Saccharomyces cerevisiae*, also functions during mitosis, where these kinases are required for bipolar spindle assembly and chromosome segregation. Although increased levels of two of the human Aurora-related kinases have been found in several human cancers^{15–18}, overexpression of one of them induces transformation of cultured cells^{15,16}.

Because overexpression of Eg2 in *Xenopus* oocytes markedly accelerated *c-mos* translation, we determined whether Eg2 could be responsible for the early phosphorylation of CPEB at Ser 174. Although baculovirus-expressed Eg2 phosphorylated recombinant wild-type CPEB *in vitro*, it did not phosphorylate CPEB-AA (Fig. 4a), suggesting that the regulatory Ser 174 residue is a target of this kinase. To confirm that this site is phosphorylated *in vivo*, we obtained phospho-peptide maps of CPEB phosphorylated *in vitro* by Eg2 (Fig. 4b, left) and *in vivo* in response to progesterone (Fig. 4b, middle). The single CPEB phospho-peptide produced by Eg2 displays a 2D mobility similar to that of Pp1. A map of the mixed peptides from *in vitro* and *in vivo* phosphorylated proteins confirms that Eg2 phosphorylates CPEB on the same site as that which occurs *in vivo* (Fig. 4b, right). Note that the *in vivo* peptide map in Fig. 4 is somewhat different (except for Pp1) from that shown in Fig. 1. This is due to the different types of trypsin used in the two experiments

(unmodified and a modified enzyme intended to reduce partial and non-specific cleavages).

The only other known substrate for Eg2 is Eg5, a kinesin-like protein¹⁹; however, its specific site of phosphorylation has not been determined. Because co-immunoprecipitation and protein 'pull-down' experiments indicate that there is relatively tight binding between Eg2 and Eg5 (ref. 19), we tested for a similar interaction between Eg2 and CPEB. We mixed His-tagged CPEB (wild-type and mutants, Fig. 4c) with oocyte extracts and captured the resulting complexes on nickel-agarose. Western blot analysis of these complexes reveals that Eg2 avidly bound wild-type CPEB, CPEB-AA and CPEB with a C-terminal deletion (Δ C) that lacks most of the RNA binding domain²⁰ (Fig. 4d). In contrast, Eg2 barely bound CPEB with an N-terminal deletion (Δ N)²⁰, despite the fact that it contains Ser 174 (Fig. 4d, lane 4). In a reciprocal experiment, His-tagged Eg2 captured CPEB when incubated with an oocyte extract (Fig. 4d, lane 7). These data indicate that Eg2 interacts with the N-terminal portion of the CPEB. This conclusion is also supported by the observation that mature oocyte extracts phosphorylated wild-type CPEB and CPEB- Δ C, but not CPEB- Δ N (Fig. 4e). The 47K (relative molecular mass 47,000) phosphoprotein that bound wild-type CPEB and CPEB- Δ C, but not CPEB- Δ N, most probably corresponds to Eg2 (Fig. 4e). These data also show that the interaction of Eg2 with the N terminus of CPEB is necessary for Ser 174 phosphorylation. CPEB- Δ N, which does not bind Eg2, acted as a dominant-negative mutation to prevent polyadenylation in mRNA-injected oocytes (Fig. 4f) and blocked maturation (data not shown), whereas CPEB- Δ C did not display any of these effects.

To conclusively show that Eg2 phosphorylates CPEB on Ser 174, the kinase was immunodepleted from a mature oocyte extract (Fig. 4g). This procedure prevented the phosphorylation of an ovalbumin-linked peptide containing residues 163–181 of CPEB, and hence Ser 174 (Fig. 4g, compare lanes 2, 4 and 6). The CPEB-derived Eg2 substrate peptide, which is an effective competitor of CPEB phosphorylation by Eg2 *in vitro* (data not shown), delayed progesterone-induced maturation when injected into oocytes, as expected of a competitive inhibitor of CPEB phosphorylation (Fig. 4h, pepWT). A non-phosphorylatable variant of this peptide, which contains the double alanine substitutions noted previously (pepAA), markedly inhibited maturation in injected oocytes, presumably because it irreversibly bound Eg2. Thus, by all criteria, Eg2 is the kinase that phosphorylates CPEB on the critical Ser 174.

These results show that Eg2, which is activated within 30 min after the addition of progesterone to oocytes¹², phosphorylates CPEB on Ser 174, and this is necessary and sufficient to trigger the cytoplasmic polyadenylation and subsequent translational recruitment of *c-mos* mRNA. Although the mechanism by which CPEB phosphorylation activates polyadenylation is unknown, it may enhance the binding of cleavage and polyadenylation specificity factor (CPSF) to the AAUAAA hexanucleotide^{21,22}, which in turn probably recruits poly(A) polymerase. Once Mos is synthesized and MAPK and maturation-promoting factor are activated, a second round of CPEB phosphorylation occurs^{5,7,8}, which may target CPEB for degradation⁵ or regulate its association with other proteins such as maskin²³. □

Methods

In vitro phosphorylation assay

Extracts were made by homogenizing five oocytes in 100 μ l of H1 kinase buffer (80 mM sodium β -glycerophosphate, 20 mM EGTA, 15 mM MgCl₂, 50 mM Na₂VO₄) with protease inhibitors (10 μ g ml⁻¹ each of leupeptin, pepstatin and chymostatin) and centrifuging (1,500g) for 5 min at 4 °C. Kinase assays were performed in 30 μ l of buffer containing 20 mM Tris pH 7.7, 10 mM MgCl₂, 50 mM KCl, 1 mM dithiothreitol and 30 μ M [γ -³²P]ATP (0.16 mCi ml⁻¹) with 10 μ l of oocyte extract or recombinant His-tagged Eg2 (ref. 12), as indicated in the figure legends. The peptide LRSSRLDSRLDSRSSC, corresponding to residues 167–183 of CPEB plus a C-terminal cysteine, was coupled to maleimide-activated ovalbumin (Pierce). Recombinant CPEB was obtained and purified as described⁶.

Construction of CPEB mutants

Point mutations were made with a Chamaleon mutagenesis kit (Stratagene). The selection primer was located in plasmid pMyc-CPEB⁵, 5'-CCTCGAGGGCGGGCCCGTACCC-AATTCGCC-3', and changed a *KpnI* site to a *SrfI* site. This primer was used in conjunction with the following mutation primers to create new clones: serines 174 and 180 to alanine (CPEB-AA), 5'-GCTCTCGATTGACGCCCCGGTCTATTTGGATGCTC-GCTCC-3'; and serines 174 and 180 to aspartic acid (CPEB-DD), 5'-GCTCTCGATTG-GACGATCGCTCTATTTGGATGCTGCTCC-3'. The His-tagged form of the mutant CPEBs was obtained by subcloning the *NcoI*-*BamHI* fragment of CPEB-AA and CPEB-DD into the *NcoI* and *BamHI* sites of pHIS-CPEB⁶.

Received 15 November 1999; accepted 24 January 2000.

1. Hake, L. E. & Richter, J. D. Translational regulation of maternal mRNA. *Biochim. Biophys. Acta* **1332**, M31-M38 (1997).
2. Fox, C. A., Sheets, M. D., Wahle, E. & Wickens, M. P. Polyadenylation of maternal mRNA during oocyte maturation: poly(A) addition *in vitro* requires a regulated RNA binding activity and a poly(A) polymerase. *EMBO J.* **11**, 5021-5032 (1992).
3. McGrew, L. L., Dworkin-Rastl, E., Dworkin, M. B. & Richter, J. D. Poly(A) elongation during *Xenopus* oocyte maturation is required for translational recruitment and is mediated by a short sequence element. *Gene Dev.* **3**, 803-815 (1989).
4. McGrew, L. L. & Richter, J. D. Translational control by cytoplasmic polyadenylation during *Xenopus* oocyte maturation: characterization of *cis* and *trans* elements and regulation by cyclin/MPF. *EMBO J.* **9**, 3743-3751 (1990).
5. Hake, L. E. & Richter, J. D. CPEB is a specificity factor that mediates cytoplasmic polyadenylation during *Xenopus* oocyte maturation. *Cell* **79**, 617-627 (1994).
6. Stebbins-Boaz, B., Hake, L. E. & Richter, J. D. CPEB controls the cytoplasmic polyadenylation of cyclin, Cdk2 and *c-mos* mRNAs and is necessary for oocyte maturation in *Xenopus*. *EMBO J.* **15**, 2582-2592 (1996).
7. Paris, J., Swenson, K., Piwnicka-Worms, H. & Richter, J. D. Maturation-specific polyadenylation: *in vitro* activation by p34^{cdc2} and phosphorylation of a 58-kD CPE-binding protein. *Genes Dev.* **5**, 1697-1708 (1991).
8. De Moor, C. H. & Richter, J. D. The *mos* pathway regulates cytoplasmic polyadenylation in *Xenopus* oocytes. *Mol. Cell. Biol.* **17**, 6419-6426 (1997).
9. Ballantyne, S., Daniel, D. L. Jr & Wickens, M. A dependent pathway of cytoplasmic polyadenylation reactions linked to cell cycle control by *c-mos* and CDK1 activation. *Mol. Biol. Cell* **8**, 1633-1648 (1997).
10. Katsu, Y., Minshall, N., Nagahama, Y. & Standart, N. Ca²⁺ is required for phosphorylation of clam p82/CPEB *in vitro*: implications for dual and independent roles of MAP and Cdc2 kinases. *Dev. Biol.* **209**, 186-199 (1999).
11. De Moor, C. H. & Richter, J. D. Cytoplasmic polyadenylation elements mediate masking and unmasking of cyclin B1 mRNA. *EMBO J.* **18**, 2294-2303 (1999).
12. Andersson, T. & Ruderman, J. V. The kinase Eg2 is a component of the *Xenopus* oocyte progesterone-activated signaling pathway. *EMBO J.* **17**, 5627-5637 (1998).
13. Glover, D. M., Leibowitz, M. H., McLean, D. A. & Parry, H. Mutations in aurora prevent centrosome separation leading to the formation of monopolar spindles. *Cell* **81**, 95-105 (1995).
14. Francisco, L., Wang, W. & Chan, C. S. Type 1 protein phosphatase acts in opposition to Ipl-1 protein kinase in regulating yeast chromosome segregation. *Mol. Cell. Biol.* **14**, 4731-4740 (1994).
15. Bischoff, J. R. *et al.* A homologue of *Drosophila aurora* kinase is oncogenic and amplified in human colorectal cancers. *EMBO J.* **17**, 3052-3065 (1998).
16. Zhou, H. *et al.* Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nature Genet.* **20**, 189-193 (1998).
17. Sen, S., Zhou, H. & White, R. A. A putative serine/threonine kinase encoding gene BTAK on chromosome 20q13 is amplified and overexpressed in human breast cancer cell lines. *Oncogene* **14**, 2195-2200 (1997).
18. Tatsuka, M. *et al.* Multinuclearity and increased ploidy caused by overexpression of the aurora- and Ipl-1-like midbody-associated protein mitotic kinase in human cancer cells. *Cancer Res.* **58**, 4811-4816 (1998).
19. Giet, R., Uzbekov, R., Cubizolles, F., Le Guellec, K. & Prigent, C. The *Xenopus laevis* Aurora-related protein kinase pEg2 associates with and phosphorylates the kinesin-related protein XIEg5. *J. Biol. Chem.* **274**, 15005-15013 (1999).
20. Hake, L. E., Mendez, R. & Richter, J. D. Specificity of RNA binding by CPEB: Requirement for RNA recognition motifs and a novel zinc finger. *Mol. Cell. Biol.* **18**, 685-693 (1998).
21. Bilger, A., Fox, C. A., Wahle, E. & Wickens, M. Nuclear polyadenylation factors recognize cytoplasmic polyadenylation elements. *Genes Dev.* **8**, 1106-1116 (1994).
22. Dickson, K. S., Bilger, A., Ballantyne, S. & Wickens, M. P. The cleavage and polyadenylation specificity factor in *Xenopus laevis* oocytes is a cytoplasmic factor involved in regulated polyadenylation. *Mol. Cell. Biol.* **19**, 5707-5717 (1999).
23. Stebbins-Boaz, B., Cao, Q., de Moor, C. H., Mendez, R. & Richter, J. D. Maskin is a CPEB associated factor that transiently interacts with eIF-4E. *Mol. Cell* **4**, 1017-1027 (1999).
24. Boyle, W. J., van der Geer, P. & Hunter, T. Phosphopeptide mapping and phosphoamino acid analysis by two-dimensional separation on thin-layer cellulose plates. *Methods Enzymol.* **201**, 110-149 (1991).

Acknowledgements

We thank E. J. Luna for her help with the two-dimensional separation of phosphopeptides, J. Leszyk for amino-acid sequencing, and the W. M. Keck Foundation for support of the Protein Chemistry Facility. We also thank M. Fernandez and members of the Richter laboratory for discussions and critically reading the manuscript. R.M. was supported by a Leukemia Society of America Special Fellow Award and L.E.L. was supported by a National Science Foundation Predoctoral Fellowship. This work was supported by grants from the NIH (to J.V.R. and J.D.R.).

Correspondence and requests for materials should be addressed to J.D.R. (e-mail: joel.richter@umassmed.edu).

The structure of malaria pigment β -haematin

Silvina Pagola^{*}, Peter W. Stephens^{*}, D. Scott Bohle[†], Andrew D. Kosar[†] & Sara K. Madsen[‡]

^{*} Department of Physics & Astronomy, State University of New York, Stony Brook, New York 11794-3800, USA

[†] Department of Chemistry, University of Wyoming, Laramie, Wyoming 82071-3838, USA

[‡] Present address: Department of Biologic and Materials Sciences, University of Michigan, Ann Arbor, Minnesota 48109, USA

Despite the worldwide public health impact of malaria, neither the mechanism by which the *Plasmodium* parasite detoxifies and sequesters haem, nor the action of current antimalarial drugs is well understood. The haem groups released from the digestion of the haemoglobin of infected red blood cells are aggregated into an insoluble material called haemozoin or malaria pigment. Synthetic β -haematin (Fe^{III}-protoporphyrin-IX)₂ is chemically^{1,2}, spectroscopically^{2,3} and crystallographically⁴ identical to haemozoin and is believed to consist of strands of Fe^{III}-porphyrin units, linked into a polymer by propionate oxygen-iron bonds. Here we report the crystal structure of β -haematin determined using simulated annealing techniques to analyse powder diffraction data obtained with synchrotron radiation. The molecules are linked into dimers through reciprocal iron-carboxylate bonds to one of the propionic side chains of each porphyrin, and the dimers form chains linked by hydrogen bonds in the crystal. This result has implications for understanding the action of current antimalarial drugs and possibly for the design of new therapeutic agents.

Numerous mechanisms have been proposed for the drug action of the quinoline-based antimalarials such as chloroquine⁵, but consensus has only been reached regarding its localization in the digestive vacuole of the trophozoite^{6,7}, its *in vitro* disruption of haem aggregation or detoxification^{8,9}, and its association with haemozoin¹⁰. The structure of haemozoin remains speculative despite the many spectroscopic tools used to characterize it^{2-4,11,12}. This aggregated product of haem detoxification is now believed to be a coordination polymer^{2-4,8-10,13-16}, principally because of the η^1 -carboxylate links between the iron and the propionate of an adjacent haem, and because monomeric haematin (Fe^{III}-protoporphyrin-IX

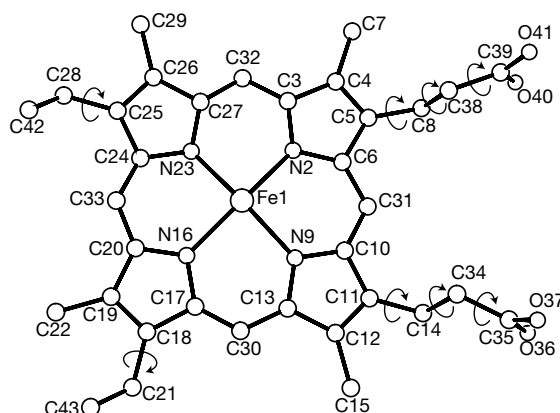


Figure 1 Representation of the Fe^{III}-protoporphyrin-IX molecule used in the structure solution (hydrogen atoms not shown). The arrows indicate torsional degrees of freedom for the vinyl and propionic acid groups. This illustration corresponds to the molecular conformation obtained in the structure solution.

(OH)) dissolves more readily than β -haematin in aqueous and coordinating solvents.

Here we describe the crystal structure of β -haematin, a well-known synthetic analogue of haemozoin. For this work to be directly relevant to malarial haemozoin, we needed to establish that the crystal structures of these two solids are identical. β -haematin is spectroscopically identical to haemozoin that is isolated from or found in malarial trophozoites^{2,3}, but this alone is not sufficient proof of solid-state structural identity. Previously, we prepared homogenous single-phase microcrystalline powders of β -haematin from the dehydrohalogenation of haemin¹⁷ and compared their powder X-ray diffraction patterns with those of whole dried parasitized red blood cells⁴. The diffraction patterns are visually identical, implying that they both correspond to the same triclinic unit cell and are isostructural at the atomic level; therefore, to solve the structure of one phase is to solve the structure of both. Two factors favour the use of synthetic β -haematin for the crystallographic analysis: a better signal-to-background ratio in the diffracted intensity, and the observation of sharper diffraction peaks. Neither of these indicates a difference in the atomic structure within each unit cell; therefore, the current results are directly transferable to natural haemozoin.

We collected high-resolution X-ray powder diffraction data at beamline X3B1 of the National Synchrotron Light Source from samples of β -haematin, both in a 1-mm capillary and in a flat-plate sample holder. The powder patterns indexed to a triclinic unit cell, $a = 12.196(2)$ Å, $b = 14.684(2)$ Å, $c = 8.040(1)$ Å, $\alpha = 90.22(1)^\circ$, $\beta = 96.80(1)^\circ$, $\gamma = 97.92(1)^\circ$; $V = 1416.0(3)$ Å³; $\rho_{\text{exp}} = 1.45(1)$ g cm⁻³; $Z = 2$, which is essentially identical to earlier results^{3,4} for other batches of β -haematin. We determined the structure by a simulated annealing algorithm, similar to described methods^{18,19}. Briefly, the chemical structure of the Fe^{III}-protoporphyrin-IX unit is known (Fig. 1), implying that the previous estimates of all bond distances and angles were good, although the eight torsion angles per molecule have to be determined. We took an initial estimate of the geometry of the porphyrin group from that of haemin²⁰. Solving a crystal structure consists of positioning the molecule(s) within the unit cell and finding the torsion angles that give the best agreement between experimental and calculated powder diffraction profiles. For any candidate structure, we define an agreement factor between the observed ($I_{\text{obs}}(2\theta)$) and calculated ($I_{\text{calc}}(2\theta)$) powder patterns as

$$S = \frac{\int d2\theta (I_{\text{obs}}(2\theta) - I_{\text{calc}}(2\theta))^2}{(\int d2\theta I_{\text{obs}}(2\theta))^2}$$

In practice, it is more efficient to compare the integrated intensities of diffraction lines, but their overlap prevents this from

being done unambiguously. We have developed a technique that compares integrated intensities A_{hkl} extracted from the data using the Le Bail algorithm²¹ with a set of B_{hkl} calculated for a candidate structure. S can be calculated very efficiently by noting that

$$S = \frac{\int d2\theta \left(\sum_{hkl} A_{hkl} f_{hkl}(2\theta) - \sum_{hkl} B_{hkl} F_{hkl}(2\theta) \right)^2}{\left(\int d2\theta \sum_{hkl} A_{hkl} f_{hkl}(2\theta) \right)^2} = \frac{\sum_{hkl, h'k'l'} (A_{hkl} - B_{hkl}) F_{hkl, h'k'l'} (A_{h'k'l'} - B_{h'k'l'})}{\left(\sum_{hkl} A_{hkl} \right)^2}$$

where $f_{hkl}(2\theta)$ is the profile of the (hkl) powder peak and the overlap

$$F_{hkl, h'k'l'} = \int d2\theta f_{hkl}(2\theta) f_{h'k'l'}(2\theta)$$

is significant for nearby peaks, and only needs to be computed once. A similar method has been described¹⁸, using the Pawley extraction method and the least-squares correlation matrix for $F_{hkl, h'k'l'}$.

Using the first 150 diffraction peaks ($2\theta < 24.82^\circ$) from the capillary data set, we made numerous simulated annealing runs in both triclinic space groups. In $P\bar{1}$, the 2 molecules in the unit cell are related by inversion symmetry, and so there are 14 parameters (3 coordinates for the centre of the molecule, 3 Euler angles to specify its orientation and 8 torsion angles). All of the structures with $S < 0.02$ in space group $P\bar{1}$ were essentially identical. In each of these cases, one of the carboxylic oxygen atoms was in the range of 1.90–2.49 Å from the iron atom, and the oxygen–oxygen distances of the other propionate groups were 3.25–3.97 Å. In every one of these solutions, the molecules were paired into dimers by two reciprocal iron–oxygen bonds as shown in Fig. 2. The annealing schedule often could not reach such a low value of S , but these metastable solutions never showed suitable bonding neighbours for the propionic oxygens. Simulated annealing in $P1$ (2 independent molecules, 25 geometrical parameters) also found a candidate structure equivalent to the $P\bar{1}$ structure described above and never found anything with a lower value of S .

A Rietveld refinement of the structure was performed using GSAS²² on the flat-plate data set. We kept the geometry of the iron–prophyrin ring as a rigid body in the Rietveld refinement, with the high-spin iron ion located 0.47 Å above the plane of the porphyrin ring. Soft bond distance constraints were included for the refinement of the positions and bond angles of the carbon and oxygen atoms of the vinyl and propionic groups. The best refinement obtained has $R_{\text{wp}} = 6.37\%$, $R_I = 3.0\%$ and $\chi^2 = 4.2$. Correspondent values in the Le Bail fit were $R_{\text{wp}} = 4.13\%$ and $\chi^2 = 1.9$. Observed and

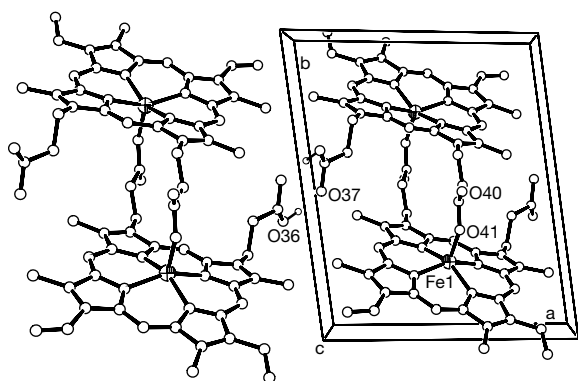


Figure 2 Two unit cells of the crystal structure of β -haematin, viewed along the [001] direction. Formation of dimers occurs through Fe1–O41 bonds, whereas dimers are linked by hydrogen bonds through O36 and O37. All other hydrogens are omitted for clarity.

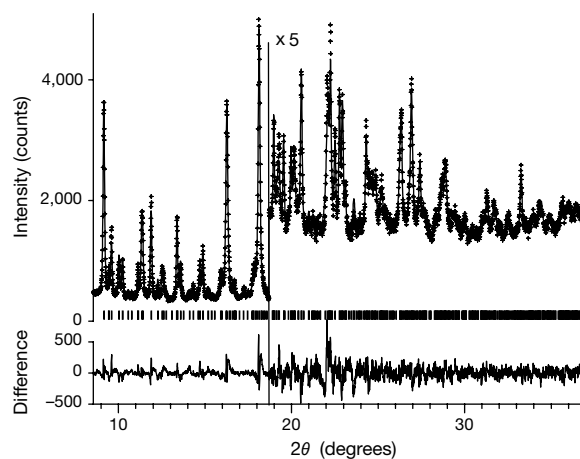


Figure 3 Observed (crosses) and calculated (solid line) X-ray powder diffraction patterns of β -haematin, using X-rays of wavelength 1.16192 Å. The lower trace shows the difference curve.

calculated diffraction patterns are shown in Fig. 3. Given the limitations of the powder diffraction technique on this sample and the size of the molecule (43 non-hydrogen atoms), the atomic coordinates are not as accurately determined as from a typical single-crystal experiment; however, the overall result of the molecular bonding geometry is unequivocally established by this work.

After the Rietveld refinement, there was not significant structure in the Fourier difference map, which ranged between -0.095 and $+0.086 \text{ e}^- \text{ \AA}^{-3}$. As an independent test of the correctness of the connection topology of the molecules, we ran Fourier difference calculations starting with the atomic positions of only the iron–porphyrin ring. The oxygen bonded to the iron and the two carbons of the propionic side chain are immediately visible. This precludes the possibility that the *S* factor in the structure solution stage is insufficiently sensitive to the propionic chain's positions which might lead to solutions in which these fragments are incorrectly located.

The refined crystal structure is shown in Figs 2 and 4. The inversion symmetry creates the dimer pair in the unit cell, and these in turn stack in the lattice to give parallel iron porphyrin units along the [001] direction. Critical metric parameters include the Fe1–O41 bond distance, which converged to a value $1.886(2) \text{ \AA}$. Refinements were satisfactory for O36–H–O37 hydrogen bonds, in the range of $2.8 \text{ \AA}$, which is in agreement with crystallographically characterized carboxylic acid dimers. The Fe–O bond length is not significantly different from those found for high-spin ferric complexes of synthetic porphyrins, for example $1.898(4) \text{ \AA}$ (acetate) and $1.847(2) \text{ \AA}$ (phenoxide)²³. The refined atomic coordinates and selected bond distances and angles are given in the Supplementary Information.

This structure is consistent with all of the known spectroscopic^{1–3} and magnetic^{3,11} properties of β -haematin. To our knowledge²⁴,

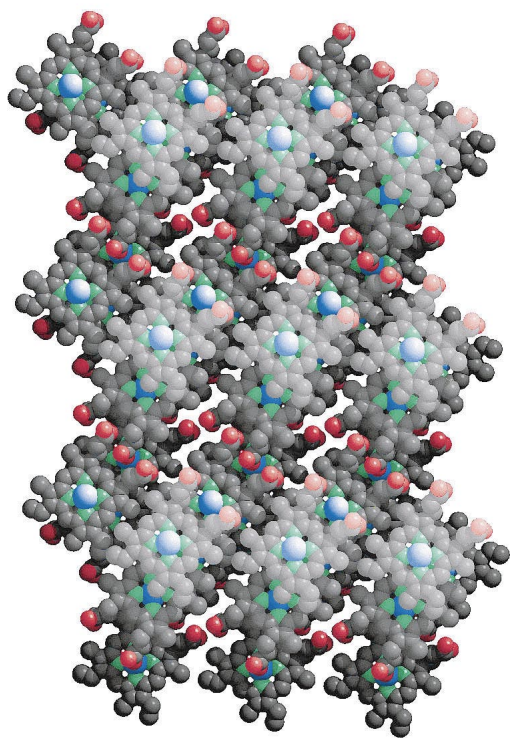


Figure 4 Packing diagram of the crystal structure of β -haematin. The [131] face, which is roughly parallel to the porphyrin rings, is in the plane of the paper. Lighter shaded atoms (C, grey; O, red; N, green; Fe, blue; H, not shown) are nearer the viewer. The chains of hydrogen-bonded dimers extend from left to right.

there has been no observation of a porphyrin dimer in which the metal atoms are reciprocally esterified by the side chains. A similar idea was first suggested²⁵ in 1949, but in that model, dimerization was proposed to be by ion pairing of a cationic six-coordinate iron to anionic propionates, which has never been experimentally confirmed.

Comparing the solubilities of haematin and β -haematin, in sufficiently alkaline solutions β -haematin will dissolve, presumably by solvation of the free carboxylate groups and rupture of the iron–carboxylate bonds. This will be true for either the dimer described here or the previously proposed polymer. On the basis of our structure, the difference in solubilities in coordinating solvents such as pyridine and dimethyl sulphoxide might be due to a chelate effect in which the kinetics of iron–carboxylate bond substitution are retarded owing to the dimeric nature of β -haematin; the rate of the back reaction following solvolysis of one Fe–OC(O) bond to the dimer will compete with the solvolysis of the second Fe–OC(O) bond to give to completely separate monomers. In the case of a polymer, each solvolysis reaction will lead to liberation of a haem. However, any discussion of solubilization of these species requires full consideration of the inhomogeneous kinetics of the crystal/solution interface, and these will be markedly influenced by crystal-lite morphology, size and the solute–solution interaction. Consequently, the observed solubility characteristics alone are insufficient evidence for the polymeric structure.

In vitro experiments have established that the quinoline anti-malarial drugs are associated with the crystallization of haemazoin¹⁰. This suggests that their therapeutic effect occurs by interference with the formation of that solid phase. Proposed mechanisms for such a drug action can be grouped into three categories: (1) direct binding of the drug to haem monomers or dimers in solution¹³, which interferes with the crystallization of haemazoin; (2) enzymatic inhibition of a protein that catalyses haemazoin crystallization⁸; and (3) chemisorption of the drug onto crystallized haemazoin, leading to inhibition of further haem aggregation²⁶.

Some models in the first category^{10,13} have proposed a haem–drug complex capping the extension of a presumed haematin polymer. As we have established that haemazoin is not a coordination polymer so that the growth of a chain cannot be stopped at a single site, these models must be re-examined. This is not merely a semantic difference, that is to substitute crystallization for polymerization. Any such model now has to explain how a $\sim 1 \text{ mM}$ intravacuolar²⁷ concentration of chloroquine could act catalytically to maintain a sufficient ($\sim 0.4 \text{ M}$) concentration of free haematin¹⁰ to disrupt the *Plasmodium* trophozoites and display the known therapeutic action.

The second category of models suggests that a protein mediates haem aggregation into haemazoin, and that the action of quinoline drugs is to inhibit this supposed haem polymerase⁸. The fact that quinoline inhibits the growth of haemazoin *in vitro* argues against this hypothesis, although the *in vitro* growth of haemazoin at physiological pH and temperature has not been unequivocally shown. Unfortunately, the isolation of proteinaceous fractions from *Plasmodium* trophozoites with unambiguous activity to form haemazoin has been controversial, and considerable uncertainty lingers over the role of enzyme in the aggregation/detoxification process^{9,14,15}.

Finally, the haematin–quinoline complex may not be covalently bonded to haemazoin¹⁰, and an evaluation of all of these mechanisms has led to the proposal of a general and as yet undetermined surface–drug interaction as the origin of quinoline's drug action²⁶. Both natural and synthetic malaria pigment crystallize as long thin needles²⁸, whose approximate dimensions are $1.0 \pm 0.1 \text{ \mu m}$, $0.6 \pm 0.1 \text{ \mu m}$ and $0.2 \pm 0.1 \text{ \mu m}$, as isolated from mature trophozoites. These remarkably uniformly sized and shaped crystals have estimated volumes and surface areas of $0.1 \text{ \mu m}^3$ and $2 \text{ \mu m}^2$

respectively. Together, these crystals feature a pair of relatively small, fast-growing faces which correspond to ~5% of the total crystal surface area. The hypothesis that antimalarial drug action of quinoline arises from surface absorption on the actively growing faces of malaria pigment crystallites is entirely consistent with the following estimate of ~8% surface area coverage for the final crystallites. (The experimentally determined density for malaria pigment, $\rho = 1.45(1) \text{ g cm}^{-3}$ from the structure and unit cell, indicates that there are 0.3 fmol of haem per crystal, and thus in a typical 80-fl red blood cell with a haemoglobin concentration of 20 mM there is enough haem to make about 10 of the final malaria pigment crystals. Absorption coverage is based on estimates of the intravacuolar chloroquine concentration of ~1 mM and chloroquine's predicted Connolly surface coverage of ~300 Å².) Clearly, for crystallites formed and growing in the young trophozoites, there will be more than sufficient chloroquine to ensure complete surface coverage, although we note that we are not aware of any experimental description of either the number or morphology of the crystallites isolated from young trophozoites. The hypothesis of a surface adsorption process limiting new haem uptake is therefore plausible, and this work provides impetus to further determine the surface structure and chemistry of malaria pigment.

Haemoglobin catabolism and β -haematin formation in the *Plasmodium* trophozoite is a highly organized process designed not just to efficiently deliver amino acids for *de novo* protein synthesis, but also to avoid the build-up and adventitious release of free haem. The organization of digestive vacuole haemoglobin peptidases is well understood^{16,29,30}. In contrast, issues regarding iron oxidation and haemozoin nucleation and growth remain poorly explained, and these must be considered as potential drug targets. Although there is evidence that quinoline surface adhesion is present, and contributes to the remarkable concentration of drug within the digestive vacuole of *Plasmodium*, there may be other interactions that also contribute to their drug action. Detailed information concerning all levels of intravacuolar haem processing is required if the biochemistry of haem detoxification is to be understood at a more meaningful level. The unexpected observation of haem dimers in the solid malaria pigment is a fundamental ingredient of any such description. □

Received 2 November 1999; accepted 7 January 2000.

1. Fitch, C. D. & Kanjanangulpan, P. The state of ferriprotoporphyrin IX in malaria pigment. *J. Biol. Chem.* **262**, 15552–15555 (1987).
2. Slater, A. F. G. *et al.* An iron–carboxylate bond links the heme units of malaria pigment. *Proc. Natl Acad. Sci. USA* **88**, 325–329 (1991).
3. Bohle, D. S. *et al.* Structural and spectroscopic studies of β -hematin (the heme coordination polymer in malaria pigment). *Am. Chem. Soc. Symp. Ser.* **572**, 497–515 (1994).
4. Bohle, D. S., Dinnebie, R. E., Madsen, S. K. & Stephens, P. W. Characterization of the products of the heme detoxification pathway in malarial late trophozoites by X-ray diffraction. *J. Biol. Chem.* **272**, 713–716 (1997).
5. Olliaro, P. L. & Goldberg, D. E. The *Plasmodium* digestive vacuole—metabolic headquarters and choice drug target. *Parasitol. Today* **11**, 294–297 (1995).
6. Yayon, A., Cabantchik, I. & Ginsburg, H. Susceptibility of human malaria parasites to chloroquine is pH dependent. *Proc. Natl Acad. Sci. USA* **82**, 2784–2788 (1985).

7. Homewood, C., Warhurst, D., Peters, W. & Baggeley, V. Lysosomes, pH and the anti-malarial action of chloroquine. *Nature* **235**, 50–52 (1972).
8. Slater, A. F. G. & Cerami, A. Inhibition by chloroquine of a novel haem polymerase enzyme activity in malaria trophozoites. *Nature* **355**, 167–169 (1992).
9. Dorn, A., Stoffel, R., Matile, H., Bubendorf, A. & Ridley, R. G. Malarial haemozoin/ β -haematin supports haem polymerization in the absence of protein. *Nature* **374**, 269–271 (1995).
10. Sullivan, D. J., Gluzman, I. Y., Russell, D. G. & Goldberg, D. E. On the molecular mechanism of chloroquine's antimalarial action. *Proc. Natl Acad. Sci. USA* **93**, 11865–11870 (1996).
11. Bohle, D. S., Debrunner, P., Jordan, P. A., Madsen, S. K. & Schulz, C. E. Aggregated heme detoxification byproducts in malarial trophozoites: β -hematin and malaria pigment have a single S = 5/2 iron environment in the bulk phase as determined by EPR and magnetic Mössbauer spectroscopy. *J. Am. Chem. Soc.* **120**, 8255–8256 (1998).
12. Adams, P. A., Berman, P. A. M., Egan, T. J., Marsh, P. J. & Silver, J. The iron environment in heme and heme-antimalarial complexes of pharmacological interest. *J. Inorg. Biochem.* **63**, 69–77 (1996).
13. Dorn, A. *et al.* An assessment of drug–haematin binding as a mechanism for inhibition of haematin polymerisation by quinoline antimalarials. *Biochem. Pharmacol.* **55**, 727–736 (1998).
14. Bendrat, K., Berger, B. J. & Cerami, A. Haem polymerization in malaria. *Nature* **378**, 138 (1995).
15. Ridley, R. G., Dorn, A., Matile, H. & Kansy, M. Haem polymerization in malaria. *Nature* **378**, 138–139 (1995).
16. Francis, S. E., Sullivan, D. J. & Goldberg, D. E. Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum*. *Annu. Rev. Microbiol.* **51**, 97–123 (1997).
17. Bohle, D. S. & Helms, J. B. Synthesis of beta-hematin by dehydrohalogenation of hemin. *Biochem. Biophys. Res. Commun.* **193**, 504–508 (1993).
18. David, W. I. F., Shankland, K. & Shankland, N. Routine determination of molecular crystal structures from powder diffraction data. *Chem. Commun.* 931–932 (1998).
19. Andreev, Y. G. & Bruce, P. G. Solving crystal structures of molecular solids without single crystals: a simulated annealing approach. *J. Chem. Soc. Dalton Trans.* 4071–4080 (1998).
20. Koenig, D. F. The structure of α -chlorohemin. *Acta Crystallogr.* **18**, 663–673 (1965).
21. LeBail, A., Duroy, H. & Fourquest, J. L. *Ab initio* structure determination of LiSbWO₆ by X-ray-powder diffraction. *Mater. Res. Bull.* **23**, 447–452 (1988).
22. Larson, A. C. & Von Dreele, R. B. *GSAS General Structure Analysis System* (Los Alamos Laboratory Report No. LA-UR-86-748, Los Alamos, 1987).
23. Bominaar, E. L. *et al.* Structural, Mossbauer, and EPR investigations on two oxidation states of a five-coordinate, high spin synthetic heme. Quantitative interpretation of zero-field parameters and large quadrupole splitting. *Inorg. Chem.* **31**, 1845–1854 (1992).
24. Scheidt, W. R. & Lee, Y. J. Recent advances in the stereochemistry of metalotetrapyrroles. *Struct. Bond.* **64**, 1–70 (1987).
25. Lemberg, R. & Legge, J. W. *Haematin Compounds and Bile Pigments* (Interscience, New York, 1949).
26. Sullivan, D. J., Matile, H., Ridley, R. G. & Goldberg, D. E. A common mechanism for blockade of heme polymerization by antimalarial quinolines. *J. Biol. Chem.* **273**, 31103–31107 (1998).
27. Slater, A. F. G. Chloroquine-mechanism of drug-action and resistance in *Plasmodium falciparum*. *Pharm. Ther.* **57**, 203–235 (1993).
28. Egan, T. J., Hempelmann, E. & Mavuso, W. W. Characterisation of synthetic beta-haematin and effects of the antimalarial drugs quinidine, halofantrine, desbutylhalofantrine and mefloquine on its formation. *J. Inorg. Biochem.* **73**, 101–107 (1999).
29. Gluzman, I. Y. *et al.* Order and specificity of the *Plasmodium falciparum* hemoglobin degradation pathway. *J. Clin. Invest.* **93**, 1602–1608 (1994).
30. Francis, S. E., Gluzman, I. Y., Oksman, A., Banerjee, D. & Goldberg, D. E. Characterization of native falcipain, an enzyme involved in *Plasmodium falciparum* hemoglobin degradation. *Mol. Biochem. Parasitol.* **83**, 189–200 (1996).

Supplementary information is available from Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

The National Synchrotron Light Source, Brookhaven National Laboratory, is supported by the US Department of Energy, Division of Chemical Sciences and Division of Materials Sciences. D.S.B. acknowledges financial support from Dreyfus Foundation and the United Nations Development Programs/World Bank/World Health Organisation Special Programme for Research and Training in Tropical Diseases. The SUNY X3 beamline at the National Synchrotron Light Source is supported by the Division of Basic Energy Sciences of the US Department of Energy.

Correspondence and requests for materials should be addressed to P.W.S. (e-mail: peter.stephens@sunysb.edu).

Why not do it by halves?

Save with 'half-area' assays and let's hear it for the grapevine in DNA preps.

Heating mantles

Gerhardt www.gerhardt.de

Heat on tap

Gerhardt's range of fully CE-compliant heating mantles for general laboratory work can be supplied in a variety of sizes, as single units, or in banks of four or six positions with either a flat hotplate or dished tubular heating element. The compact, multi-position mantle is capable of accommodating as many as six round-bottomed flasks up to 750 ml, but is only 90 cm long to make the most of valuable bench space.

Reader Service No. 100

Biochemicals and antibodies

Evotech Biologics www.evo-tech.com

A 5% discount for online purchasers

Evotech Biologics say their aim is to provide low-cost biochemicals and antibodies for apoptosis, cancer research, multi-drug resistance, signal transduction and other areas of research. Products in the current catalogue include: the apoptosis inducer Antifas, human clone IPO-4; FMK caspase inhibitors; caspase substrates; caspase enzymes; anti-granzyme B; and TNF products.

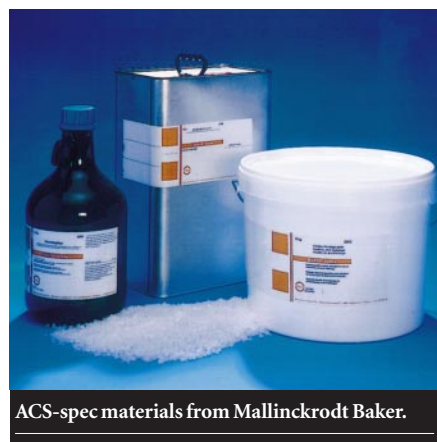
Reader Service No. 101

Specialized solvents

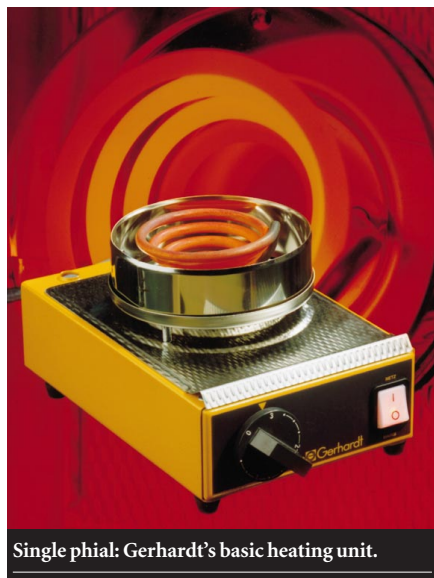
Mallinckrodt Baker www.jtbaker.nl

Solvents and alcohols for histology and cytology

The Mallinckrodt Baker broad range of alcohols includes methanol, ethanol, ethanol denatured with methanol, IPA and benzyl alcohol. Solvents in the J. T. Baker range include UltraClear, a non-toxic, non-odorous, non-carcinogenic isoparaffin xylene replacement. Xylene and toluene are also



ACS-spec materials from Mallinckrodt Baker.



Single phial: Gerhardt's basic heating unit.

available in the 'histo' grade and all are specially prepared for use in most histology lab techniques.

Reader Service No. 102

Red peroxidase substrate

Vector Laboratories www.vectorlabs.com

A new aid for immunohistochemistry

The NovaRED Substrate Kit produces a red reaction product, with about the same sensitivity as DAB. Unlike AEC, sections stained with this peroxidase substrate can be dehydrated in ethanol, cleared in organic solvents such as xylene, and permanently mounted in non-aqueous media. Vector's NovaRED Substrate Kit produces a crisp, dense reaction product which contrasts well with haematoxylin. This substrate can be used for multiple-label immunohistochemistry because it is compatible with several other peroxidase and alkaline phosphatase substrates.

Reader Service No. 103

Mini-Catalogue

Alexis www.alexis-corp.com

Available from alexis-ch@alexis-corp.com

A new 32-page catalogue, *Nitric Oxide Pathway and Oxidative Stress Research*, is available from Alexis Biochemicals — online or via mail. This comprehensive mini-catalogue features listings and detailed descriptions for nearly 300 reagents. Among the 20-plus new items are NO inhibitors, NO donors, spin-traps, antioxidants, products for the study of lipid peroxidation, and guanylyl cyclase and related products.

Reader Service No. 104

Lovastatin

Wako Bioproducts www.wakousa.com

Inhibit cholesterol biosynthesis

Lovastatin, a cholesterol-lowering agent isolated from a strain of *Aspergillus terreus*, inhibits 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase, the rate-controlling enzyme in cholesterol biosynthesis. As a pharmaceutical, lovastatin is effective in reducing total cholesterol and LDL cholesterol in heterozygous familial hypercholesterolaemia patients. Wako now include lovastatin, for biochemical use, in their bioproducts range.

Reader Service No. 105

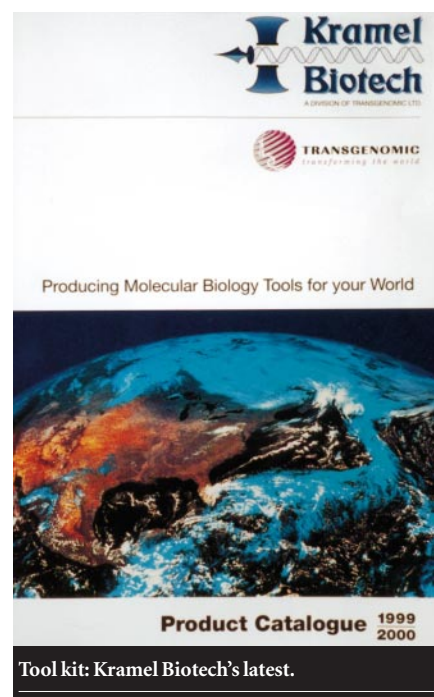
Molecular biology

Kramel Biotech www.kramel.co.uk

A new catalogue, available from kramel.biotech@lineone.net

The catalogue, *Producing Molecular Biology Tools for your World*, targeted at molecular biologists and biochemists, describes Kramel's range of life-science laboratory products. The product range includes restriction and modifying enzymes, buffers, DNA markers and products, biochemical cloning and expression vectors, SafeBox *E. coli* strains, DNA/protein purification systems, cDNA and genomic libraries, electroporation equipment and liquid handling products.

Reader Service No. 106



Tool kit: Kramel Biotech's latest.

new on the market



Half measures: Corning's economy plates.

'Half-area' assay plates

Corning Costar www.corning.com

A novel way of conserving chemical compounds and biological reagents

Costar's range of 96-well half-area assay plates, built to the proposed industry standard for critical external dimensions, are fully compatible with standard 96-well equipment. But their small well volume of 0.19 ml can help make substantial savings on rare and/or expensive reagents. Corning's tests indicate that these 96-well half-area plates can be utilized to lower reagent costs while maintaining the common 96-well footprint, with no significant reduction in sensitivity.

Reader Service No. 107

Masterpure

Epicentre Technologies Corp.

www.epicentre.com

Genomic DNA from plants in under an hour

The Masterpure Plant Leaf DNA Purification Kit is designed for the easy isolation of DNA from 35–100 mg of leaf tissue. The kit was optimized with the leaves of the grapevine *Vitis vinefera*, which presents a challenge as it has leaves high in polyphenolic compounds and polysaccharides. No enzyme digestions, CTAB or organic solvent extractions are needed in the MasterPure protocol.

Reader Service No. 108

LEF-1/TCF family antibodies

Kamiya Biomedical Company

www.kamiyabiomedical.com

A family affair

LEF-1/TCF transcription factor family members play a crucial role in WNT/Wingless signalling, a transduction cascade that directs cell differentiation. Kamiya have released three antibodies raised against full-length human LEF-1 and directed against LEF-1/TCF transcription factor family members. Polyclonal anti-LEF-1/TCF was raised in sheep and binds all four members of the LEF/TCF transcription factor family. Mono-

clonal antibody REMB6 is a mouse IgG1 that recognizes the High Mobility Group DNA binding domain of LEF-1, TCF-1 and TCF-4. Monoclonal antibody REMBI is a mouse IgG1 that recognizes the transcription activation domain of LEF-1. This antibody does not crossreact with TCF-1, TCF-3 or TCF-4.

Reader Service No. 109

AP detection

Pierce

www.piercenet.com

High sensitivity and low background in western blots

Lumi-Phos WB Chemiluminescent Substrate (Product No. 34150) is claimed to overcome the limitations posed by conventional chemiluminescent substrates for alkaline phosphatase (AP). Lumi-Phos WB Substrate provides sensitivity in the low picogram (1.2 pg) range, so mere attomoles of the target ligand can be detected. Lumi-Phos WB also produces low levels of background noise, for a high signal-to-noise ratio and a clear image. Signal generation is immediate, so there is no need to wait 15 or 30 minutes for a measurable signal.

Reader Service No. 110

Software

Galactic Industries Corporation

www.galactic.com

Download My Instrument

My Instrument is an open standard specification for developing 32-bit instrument control and data acquisition software for all types of laboratory analytical instrumentation. *My Instrument* technology provides developers with specifications for developing a reusable instrument control application with a set of common interfaces. The *My Instrument* application programming interface is based on Microsoft's ActiveX Component Object Model (COM) and is specifically designed for current and future 32-bit Windows operating systems (Windows 95 and Windows NT). *My Instrument*-compliant modules allow for setting up the hardware, setting parameters and initiating data acquisition, and displaying and storing the collected data.

Reader Service No. 111

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card.

ADVERTISEMENTS

Look no further

for the best value in

Oligo Synthesis

Competitive pricing
Routine 24 hour dispatch • Every Oligo QC verified by PAGE
Rigorous batch QC by PCR and Mass Spec
Web ordering and tracking • Local Freephone numbers
A great deal in any language!!

SIGMA (+44) (0) 1223 839000 • fax (+44) (0) 1223 839200
GNDYS email: info@sigma-genosys.co.uk • www.sigma-genosys.co.uk

READER ENQUIRY NO. 33

Diagnostic Reagents

Pefachrome®-Series: chromogenic and fluorigenic substrates

Pefabloc®-Series: serine proteinase inhibitors

Purified snake venom proteins (e.g. Protac®, Ecarin, Batroxobin)

PENTAPHARM LTD
CH 4002 Basel Fax: +41 61 319 96 19
<http://www.pentapharm.ch>

READER ENQUIRY NO. 16

L-012 SODIUM SALT

CHEMILUMINESCENCE PROBE

- CHL Intensity up to 100 times greater than MCLA & Luminol
- Works on cells & whole blood
- Ideal for Precious Cells & High Throughput Assays

Wako Wako USA WAKO GmbH
877-714-1921 (49)02131 3110
www.wako-chem.co.jp

READER ENQUIRY NO. 34

BACs
YACs
PACs
cDNAs

Research Genetics provides clones and colony membranes for a number of human, mouse, and rat libraries as well as a custom screening service for the libraries.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

READER ENQUIRY NO. 4

CULTURE BUSTING

A few interdisciplinary research case studies

p313

TRAINING FOR THE FUTURE

One forward-thinking foundation's approach

p315

BUILDING BRIDGES

A personal account

p316

WEB LINKS

For more information, see web addresses

p316

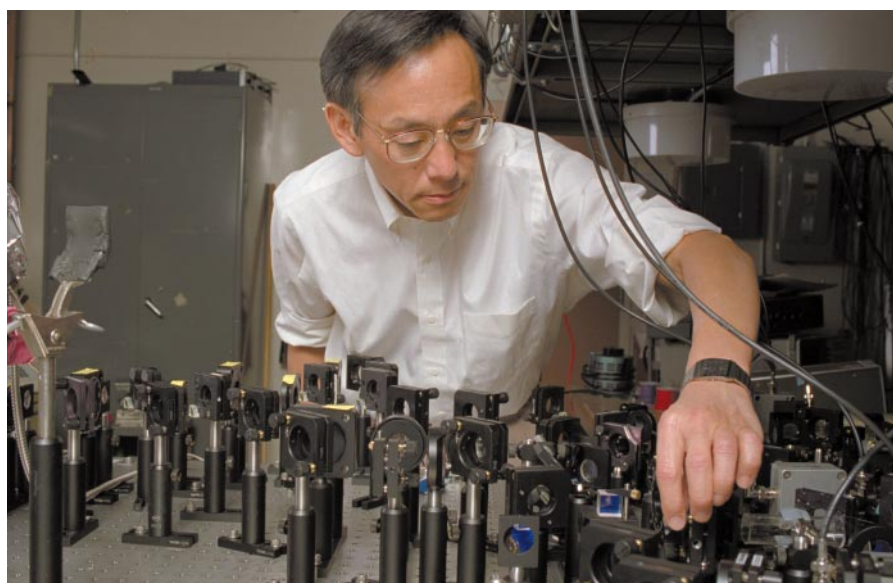
Pushing the frontiers of interdisciplinary research: an idea whose time has come

Opportunities for research at the interface between biology and the more quantitative sciences are growing fast. Diane Gershon assesses the field.

Like it or not, biological research is becoming more quantitative, the tools of the trade are becoming ever more sophisticated, and exciting discoveries are increasingly being made at the intersection of what once were disparate disciplines. Awareness is growing in biological circles of the need to foster interaction between biologists and chemists, computer scientists, engineers and physicists, who seem only too willing to apply their quantitative skills towards biological problems. Of course there is also a more practical consideration: there is simply more money in biology, and these groups know that. Whatever the reasons, institutions (and this feature is not meant to be exhaustive) with strengths in these areas are breaking down traditional cultural and bureaucratic barriers and building centres of excellence that are not dominated by any one department or discipline, and where there is a heavy emphasis on technology development.

No one is suggesting that a researcher working at the interface should become a jack of all trades. But researchers need to adopt enough of a common language and culture to communicate effectively with people from other disciplines. The task of how best to fashion these programmes presents a considerable organizational challenge for which there is no one-size-fits-all solution. One thing is certain: the approaches being tried are a marked departure from the traditional cottage-industry approach to biological research, "where biologists scraped away adding their grain of sand to the accumulated knowledge", says Edward Penhoet, dean of the School of Public Health at the University of California, Berkeley (UC Berkeley).

Stanford University and UC Berkeley have both launched major faculty-led interdisciplinary initiatives, which are broad in scope and draw on key internal strengths. They are also taking greater advantage of proximity to facilities such as the Department of Energy's Stanford Linear Accelerator (SLAC) and



Steve Chu, champion of Stanford's Bio-X initiative, in his laser lab.

Lawrence Berkeley National Laboratory as a way to boost research capabilities in the important area of structural biology — and, with Lawrence Berkeley, in genomics too.

Bio-X: a new frontier for Stanford

It is at these boundaries, or 'interdisciplinary frontiers', where exciting things are happening. Stanford's Bio-X project will be "promoting that to the hilt", says James A. Spudich, professor of biochemistry at Stanford. Bio-X is not just good for clinical medicine, he says; it will keep Stanford as a place of pre-eminence in all these areas. It also means new technology developments will continue to abound and be spun off into the biotech industry and Silicon Valley.

In terms of construction projects, SLAC was bigger, says Channing R. Robertson, professor of biochemical engineering at Stanford. However, no initiative has attempted to span three schools at once — in this case engineering, medicine, and humanities and sciences — and house some of their staff under one roof. Spudich and Steve Chu, professor of physics and applied physics at Stanford, began

hatching the idea for Bio-X almost two years ago. To them, it seemed a natural for Stanford. "It's so obvious that it didn't take more than a couple of months to convince all the leaders of the university that we had to do this," says Spudich. Biology, chemistry, engineering and physics are already strong disciplines individually, he says, and even before Bio-X there were pockets of activity in bioengineering, biocomputation, imaging and functional genomics. The university's layout is on their side too: few others have the medical school so close to the rest of the campus, although this advantage has not been used to the extent it should have been, he says.

For the past year, Spudich, Chu, Robertson and William C. Mobley, chair of the department of neurology and neurological sciences, have been hashing out the details. The group identified several hundred faculty working on problems relating to the life sciences, says Robertson, so it became clear early on that Stanford would never be able to build a facility to house everyone. The decision was made, therefore, to build a centrally located facility that would become a hub for the Bio-X

Diane Gershon is Assistant Editor, New Technology, at Nature Medicine.

programme. The building would be used to facilitate and sustain interactions between people in the building and, just as importantly, the broader academic community.

The university quickly raised \$150 million through a single donation from computer visionary and entrepreneur Jim Clark, but plans to raise considerably more. Clark is the ex-Stanford faculty member who founded Silicon Graphics and co-founded Netscape. The Clark Center, as it will be known for short, will house staff from many departments who share common intellectual interests. Spudich will be its first director. "The idea is not to form a department and have it insulate itself from the rest of the campus in the usual way of departments," says Chu.

Construction of the new building, which will cost \$120 million, is expected to begin this autumn and be completed within two years. It will house 45 faculty members, two-thirds of whom will be existing Stanford faculty, with the rest still to be recruited. There will also be some uncommitted 'hotel spaces' so everyone on campus has a potential claim.

The group says it is close to identifying key research themes and which faculty members will move to the new building. Personality is important in this regard, says Spudich, because the project needs people who want to mix and share with researchers from other disciplines. He says it is also important to choose people with strong departmental ties, so that the flow of information between the hub and outlying departments is a two-way street.

The Bio-X team (as is the case at UC Berkeley) is also devising ways to encourage staff and students who will not be based in the Clark Center to become involved. One way is to rethink the way science buildings are designed. Most are functional but uninviting, with long corridors and few places for interaction. Classrooms or lecture halls are typically located on the ground floor, with labs and offices stacked above. "No one goes upstairs unless they have to," says Robertson. The plan, therefore, is to commission a design that is both open and inviting, where facilities are distributed throughout. Another way is to make reasons for people to go there, whether it is to eat, to take classes or to use shared core facilities.

"What's really interesting is the possibility that we have no clue what will go on in the Clark Center. That's the point. Much of what we think works is this random collision that has a physics person talking to somebody interested in Alzheimer's," says Mobley. What the Clark Center is doing is increasing the likelihood that those collisions will occur.

Berkeley's answer

Like Bio-X, the health-sciences initiative at UC Berkeley has been a grass-roots effort, which will promote cross-disciplinary research in health-related areas and will involve researchers from the biological and

physical sciences as well as public health and the social sciences. "Our programme simply provides a little structure and recognition for what was already happening organically," says Penhoet. "I didn't start the revolution on campus; I came back to ground which had already been fertilized." Penhoet began his career on the faculty at Berkeley but left in 1981 to co-found Chiron, now one of the world's largest biotechnology companies. He returned after 17 years with the company.

"We already have 350–400 faculty members on campus who do biology, so we don't really need that many more. What we need is better organization," says Robert T. Tjian, professor of molecular and cell biology at Berkeley, Howard Hughes Medical Institute investigator and chair of the Chancellor's Advisory Council on Biology. Tjian, along with Penhoet, has been instrumental in pushing the health-sciences initiative forward. The initiative is a major undertaking in terms of both infrastructure and cost, he says, even for a large university like Berkeley.

Although many things about the initiative are still fluid, two new facilities are planned for the campus on sites of existing buildings. The university must raise \$300 million for the buildings alone, about half of which has already been raised, says Penhoet, including \$50 million from an anonymous donor. A number of research areas are being targeted, "picked not only because we think they're a major area of future interest but also because we have strengths there," says Tjian. These include bioengineering, cancer, microbial biology, neuroscience, post-genomics biology (including bioinformatics) and structural biology.

Neuroscience is a natural for this kind of initiative, says Corey S. Goodman, professor and director of the Wills Neuroscience Institute, in terms of understanding how the brain

works, how it is put together, how it learns and remembers, and the clinical implications of all the ways it goes wrong. "It's not clear that any of the traditional disciplines alone are going to solve that problem," he says. But neuroscience tends to be fragmented, with cognitive and behavioural scientists off in psychology departments at one end of the field and researchers studying genes and molecules at the other. Goodman says the first step has been to build a neuroscience programme that fuses the two ends together, where staff and students interact at all levels. "The different ends of the field are either getting fuelled by, or are waiting for, major technological advances that are really going to allow for paradigm shifts," he adds. The development of a number of 'technological cores' to promote interactions among people from different disciplines was already under way when plans for the broader health-sciences initiative surfaced.

The new Brain Imaging Center is one such core. The centre will house a newly purchased, state-of-the-art 4 T magnetic resonance imaging (MRI) machine, one of only a handful in the world, says Goodman. Typically, MRI machines are 'turnkey' machines located in radiology departments in hospitals and medical schools. They are patient-based and paid for on a recharge basis, he says, and so access to the machines for research purposes can be limited. Goodman says the decision to put the new 4 T machine in a basic research unit, which has nothing to do with recharges, was probably instrumental in bringing MDs Robert T. Knight of UC Davis and Mark T. D'Esposito of the University of Pennsylvania to Berkeley. Unlike their previous universities, Berkeley doesn't have a medical school (although it does collaborate with the medical school at UC San Francisco).

The next order of business in this 'lab without walls' health-sciences initiative will be to characterize further the research interests of the 350 faculty who have so far expressed an interest in being part of the initiative. An informal network exists, but there are plans to establish an interactive website to link this virtual community. Members of the steering committee are still kicking around ideas and say it is too early to identify who will eventually move into the new buildings. "As soon as we start putting names on labs then everyone will start worrying about their turf and will stop thinking big," says Goodman.

A few good fellows

On a rather smaller scale, the new Harvard Center for Genomics Research is one of several interdisciplinary science initiatives to be launched by the university, which has committed \$150–200 million to it over the next five years. Functional genomics was one of several 'critical and emerging' fields singled out by the university that draw on multidisciplinary expertise. A Center for Imaging

A number of research areas have been targeted, picked not only because we think they're a major area of future interest but also because we have strengths there



Robert Tjian



The different ends of the [neuroscience] field are either getting fuelled by, or are waiting for, major technological advances that are really going to allow for paradigm shifts **Corey Goodman**

and Mesoscale Structures is also taking shape, and others are planned.

The genomics centre, which will eventually be housed in a new 5,574-square-metre building slated for completion in autumn 2001, will provide "a non-traditional environment for non-traditional thinkers", says Dari Shalon, the centre's director. What sets this centre apart from the rest is that it won't be staffed by faculty. Instead, it will be constructed around 10 research fellows who are between a PhD and junior faculty level. This will provide "built-in renewal every three or four years", says Shalon, as these will be non-tenure-track appointments.

The fellows will not be expected to teach or to apply for grants during most of their three-year term of employment (which may be stretched to five years, depending on mutual interests). This will free them up to do research full-time, alongside people from other disciplines, and fully funded by Harvard, says Shalon.

Fellows will be selected on the basis of their ability to work with colleagues in other fields; a number of strategic areas have been identified. These include bioinformatics, biological imaging, evolutionary analysis, genome manipulation, microarray technologies, protein analysis and systems analysis. Shalon says he is looking for people who take a global view of biology and who are willing and interested in developing novel tools. He says these tend to be people with fairly diverse educational backgrounds. "They tend to self-select," he says.

Each fellow will have at least four advisers from three or four different departments. Shalon hopes these fellows will institutionalize the interdisciplinary links between departments. The use of mentors is just one way the centre aims to reach out to the Harvard community. Faculty will be also be encouraged to use the centre's microarray or bioinformatics core facilities, for example, or to send a graduate student there for training. They might also be asked to teach a course.

Shalon has himself had a varied career. He has a background in engineering, and was involved in the development of DNA-chip technology at Stanford, but also gained

industry experience by forming Synteni, which was later sold to Incyte Pharmaceuticals. He also did a stint in the Israeli army. Shalon says he was enticed back to head the new centre last summer because it "allows me to focus on issues that are several years out on the horizon". Although he has great respect for those in industry, he believes the real breakthroughs will come from people looking beyond quarterly earnings reports.

The Novartis experiment

That fact hasn't escaped Novartis and the Novartis Research Foundation. The foundation has committed \$250 million over the next 10 years to develop the new Genomics Institute of the Novartis Research Foundation in La Jolla, California. New combinatorial chemistry methods, high-throughput cell-based and mouse phenotype screening technologies, mRNA/protein profiling and structural genomics tools, as well as bioinformatics and computational tools, are allowing researchers to design, execute and analyse experiments, not one at a time but hundreds to millions at a time. This would

have been impossible 10 years ago, says Peter Schultz, director of the institute.

The aim is to bring all these tools under one roof "and use them synergistically to do biological discovery", he says. Scientists involved in technology development will work side by side with small groups of discovery scientists focused on specific biological problems. Roughly half will work on tool development and the rest on discovery research, with crossover. The ability to bring such a broad platform of technologies under one roof is beyond the scope of most academic institutions, says Schultz. On the other hand, the emphasis in companies is primarily on product development and so discovery research often takes a back seat. The foundation recognized this and was willing to back a different kind of venture. If there was a model that best approximated this approach, says Schultz, it would be the old Bell Labs. He says the Bell Labs, with their substantial technological infrastructure and highly collaborative and interactive environment, proved an incredibly fertile ground for innovation and discovery in physics and materials. "We'd like to basically do the same thing in biology."

The new institute benefits from its ties with Novartis by gaining access to interesting molecules and disease models, says Schultz. In return, through the foundation, Novartis can obtain rights to developments in the institute, or it can pass up those rights. Although the institute currently occupies temporary space, construction began this month on a new building across from Scripps, with which it maintains close links, and is slated for completion in 18 months. Schultz says the institute's workforce is growing rapidly and he expects the number of people there to more than double from the present 90 over the next two years. ■

Changing the face of training for science at the interface

In the mid-1990s, the Burroughs Wellcome Fund (BWF) identified the funding of training programmes in interdisciplinary research as a vacant niche where it could make a significant impact. The BWF-sponsored interdisciplinary training programme, which had its first call for proposals in 1996, seeks to support innovative training programmes at the interface between the quantitative physical sciences and biology that address the language and cultural barriers unique to interdisciplinary research.

One of the few stipulations is that students and postdocs have their initial training in the quantitative sciences (physics, chemistry, computer science or maths) and seek to apply

those skills to biological problems. The BWF is not "looking for some magic formula", says Nancy S. Sung, a programme officer at the fund. She also says it has

no bias towards a particular scientific area and has intentionally set out to fund different models, in the hope of finding out what works and what doesn't.

When BWF announced its call for proposals in 1996, "it was natural for us to apply", says José N. Onuchic, professor of physics at the University of California, San Diego (UCSD), and co-director of the La Jolla Interfaces in



► Science (LJIS) programme, one of four training programmes backed by BWF in its first round of awards. LJIS is a joint effort between UCSD, the Scripps Research Institute, the Salk Institute and the San Diego Supercomputer Center. Caltech, Rockefeller University and the Program in Maths and Molecular Biology were the other grant winners in the 1996 funding round.

Two other training programmes were funded in 1998, one in genomics, run jointly by Johns Hopkins University, Baltimore, and The Institute for Genomic Research, Rockville, Maryland, and the other in neuroscience at Brown University, Rhode Island. The submission deadline for a third round of awards, to be made this autumn, is 10 April.

Typically, the BWF programme provides \$500,000 a year for five years, although this year smaller institutions can opt to come in at a lower funding level. The money can be used to support student stipends, seminar series, travel to conferences or equipment purchases, for example, but not to support faculty.

"BWF's vision for proposing such a programme was instrumental in [LJIS] achieving our goals," says Onuchic. The LJIS programme aims to provide the working environment, mentoring and financial support that will allow pre- and postdoctoral fellows to apply their quantitative skills to biological problems, and to foster research at the interface, promoting collaborations between groups on both sides.

One unique aspect of the programme is the requirement for dual mentorship, whereby each prospective fellow has to choose a mentor from both the physical and biological sciences. LJIS's steering committee believes this is key to the project's success, as the dual mentors are expected to be active participants in the programme and to bring different perspectives to a problem. At the postdoc level, they are also expected to provide matching funds to support a trainee. In return, they gain access to high-calibre scientists working at the interface on fairly risky projects they might otherwise be unwilling to support from their own funds.

Rockefeller University, New York, also

received backing from BWF for its interdisciplinary training programme. It is perhaps a little unusual, as universities go, in that it has no formal departments. Instead, it exists as a collection of 75 laboratories, all reporting directly to the president. Nevertheless, in the mid-1990s, the university created two interdisciplinary centres, the Center for Studies in Physics and Biology and the Pels Family Center for Biochemistry and Structural Biology. In part, this was "to help provide a mechanism for taking advantage of funding opportunities that required a cohesive structure, as well as to promote the development of infrastructure", says Stephen K. Burley, investigator in the Howard Hughes Medical Institute at Rockefeller University and director of the Pels Center. Taken

together, the centres encompass about one-third of Rockefeller faculty members.

Burley says that when the university was looking for support for its training programme, "there was nothing else available, because the programme was so young. Our success hinged on someone like BWF being willing to gamble on a nascent programme and help develop it." The programme supports graduate students and postdocs for up to two years, while trainees have access to a range of facilities for interdisciplinary studies. The Pels centre, for example, boasts a mini-supercomputer facility. Other facilities include a biophysical-tools teaching lab and an X-ray beamline for protein crystallography being built at the Brookhaven National Laboratory. ■

Crossing the divide between theory and practice

When Dan Rokhsar took an undergraduate molecular biology course in 1980, the subject made little sense to him because, he says, it was hard to see how the facts were interrelated. By contrast, physics seemed very orderly and he liked that order. He opted for a career in theoretical physics, and during his first five years on the faculty of the physics department at the University of California at Berkeley, he worked on mathematical modelling of the behaviour of materials. Five or so years ago, Rokhsar started to develop an interest in biology. He attended graduate courses in biology in his spare time, although his department was largely unaware of this.

Rokhsar found that much had changed since his original foray into biology and was quickly smitten, although at first "it was pretty clear that I didn't even know how to phrase the questions properly", he says. He spent a

summer working in a biology lab, again without officially informing his department, to gain experimental experience.

"In physics you can be a theorist, and that's a perfectly respectable way to make a living," he says. "In biology, you don't have that kind of division of labour as much: everybody's a theorist, but your theories are typically much less mathematical." Of the few colleagues he did tell, Rokhsar says that at least one suggested that, rather than becoming a student again, he should go to a biology professor and say: "I'm a physics professor. Tell me your difficult problems and let me help you solve them." That doesn't work, says Rokhsar, because if you don't understand the context, nothing productive will come of it.

Rokhsar, now a professor of physics and biophysics at Berkeley, says that, having decided that he wanted to switch his research focus, he looked at three areas of study—the mechanisms of protein folding, modelling in neuroscience, and devising faster ways of analysing the vast amounts of data generated from DNA microarrays. "One of the advantages of being a theorist is that you can do a lot of different things. What they all have in common is that they use mathematical models to describe complicated systems, even though the systems themselves can be very different," he says.

Rokhsar's connections with biology have been further consolidated in that he now sits on the Chancellor's Advisory Council on Biology, which oversees the entire biology programme on campus. He says the group has been instrumental in bringing in new faculty, some through joint appointments, working at the interface between the physical sciences and biology. The biophysics group itself is expected to grow by at least three people over the next five years or so. ■

Web watch

● **Stanford University's Bio-X initiative**
<http://biochem.stanford.edu/biox/>

● **UC Berkeley**
<http://www.berkeley.edu>

● **Harvard's Center for Genomics Research**
<http://www.cgr.harvard.edu/>

● **Genomics Institute of the Novartis Research Foundation**
<http://www.nfire.org/>

● **Rockefeller University: Center for Biochemistry and Structural Biology**
<http://www.rockefeller.edu/graduate/cenbio.htm>
Director, Stephen K. Burley
Center for Studies in Physics and Biology
<http://www.rockefeller.edu/graduate/cenphys.htm>
Director, Mitchell J. Feigenbaum

● **LJIS**
<http://ljis.ucsd.edu>
Co-director, Jose N. Onuchic

(jonuchic@ucsd.edu)
Co-director, Elizabeth Getzoff
(edg@scripps.edu)

● **The Burroughs Wellcome Fund**
<http://www.bwfund.org>
Program officer, Nancy S. Sung

● **The Sloan Foundation**
<http://www.sloan.org/bios/teitel.htm>
Program officer, computational molecular biology, Michael S. Teitelbaum